

## SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: \_\_\_\_\_ Examiner #: \_\_\_\_\_ Date: \_\_\_\_\_  
 Art Unit: \_\_\_\_\_ Phone Number 30 \_\_\_\_\_ Serial Number: \_\_\_\_\_  
 Mail Box and Bldg/Room Location: \_\_\_\_\_ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.  
 \*\*\*\*\*

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc. if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: \_\_\_\_\_

Inventors (please provide full names): \_\_\_\_\_

Earliest Priority Filing Date: \_\_\_\_\_

*\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

## STAFF USE ONLY

Searcher: Toby Port  
 Searcher Phone #: 308-3534  
 Searcher Location: \_\_\_\_\_  
 Date Searcher Picked Up: 8/17  
 Date Completed: 8/20  
 Searcher Prep & Review Time: 10  
 Clerical Prep Time: \_\_\_\_\_  
 Online Time: 10

## Type of Search

NA Sequence (#) 4  
 AA Sequence (#) 2  
 Structure (#) \_\_\_\_\_  
 Bibliographic \_\_\_\_\_  
 Litigation \_\_\_\_\_  
 Fulltext \_\_\_\_\_  
 Patent Family \_\_\_\_\_  
 Other \_\_\_\_\_

## Vendors and cost where applicable

STN \_\_\_\_\_  
 Dialog \_\_\_\_\_  
 Questel/Orbi: \_\_\_\_\_  
 Dr. Link \_\_\_\_\_  
 Lexis/Nexis \_\_\_\_\_  
 Sequence Systems 05  
 WWW/Internet \_\_\_\_\_  
 Other (specify) \_\_\_\_\_

\_\_\_\_\_



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 18, 2001, 22:49:18 ; Search time 3118.05 Seconds  
(without alignments)  
6163.356 Million cell updates/sec

Title: US-09-284-320-56

Perfect score: 2033

Sequence: 1 gagtcgagcgcgtaccctc.....gttaatgatgtaccacaaac 2033

Scoring table: OLIGO\_NDC  
Gapop 60.0 , Gapext 60.0

Searched: 1022815 segs, 4726426750 residues

Word size : 10

Total number of hits satisfying chosen parameters: 12737487

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

EST:\*

```
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_est11:*
64: em_est12:*
65: em_est13:*
66: em_est14:*
67: em_est15:*
68: em_est16:*
69: em_est17:*
70: em_est18:*
71: em_est19:*
72: em_est20:*
73: em_est21:*
74: em_est22:*
75: em_est23:*
76: em_est24:*
77: em_est25:*
78: em_est26:*
79: em_est27:*
80: em_est28:*
81: em_est29:*
82: em_est30:*
83: em_est31:*
84: em_est32:*
85: em_est33:*
86: em_est34:*
87: em_est35:*
88: em_est36:*
89: em_est37:*
90: em_est38:*
91: em_est39:*
92: em_est40:*
93: em_est41:*
94: em_est42:*
95: em_est43:*
96: em_est44:*
97: em_est45:*
98: em_est46:*
99: em_est47:*
100: em_est48:*
101: em_est49:*
102: em_est50:*
103: em_est51:*
104: em_est52:*
105: em_est53:*
106: em_est54:*
107: em_est55:*
108: em_est56:*
109: em_est57:*
110: em_est58:*
111: em_est59:*
112: em_est60:*
113: em_est61:*
114: em_est62:*
115: em_est63:*
116: em_est64:*
```

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

```

---

```

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES									
Result No.	Score	Query Match	Length	DB	ID	Description			
1	652	32.1	681	154	BG530418	BG530418 602559169	C	68	415
2	642	31.6	665	175	BG288867	BG288867 602838308	C	69	415
3	609	30.0	852	154	BG501143	BG501143 602463582	C	70	415
4	602	29.6	947	141	BE897106	BE897106 601439663	C	71	413
5	591	29.1	619	113	AM249865	AM249865 2821628.3	C	72	411
6	584	28.7	586	16	AI091273	AI091273 0w62603.x	C	73	410
7	572	28.3	572	168	BF726753	BF726753 by11d12.y	C	74	409
8	554	27.3	954	172	BF696051	BF696051 602269980	C	75	408
9	551	27.1	564	141	BB894114	BB894114 601438242	C	76	407
10	544	26.8	546	12	AA854119	AA854119 aj17b08.s	C	77	406
11	531	26.1	538	24	AI741720	AI741720 wq22810.x	C	78	405
12	529	26.0	966	153	BG403007	BG403007 602418818	C	79	405
13	521	25.6	521	7	AA427616	AA427616 zw30c10.s	C	80	403
14	518	25.5	648	122	AM968992	AM968992 EST81069	C	81	403
15	516	25.4	516	143	BF060652	BF060652 7159a04.x	C	82	401
16	508	25.0	508	23	AI678997	AI678997 lu60f08.x	C	83	401
17	507	24.9	538	15	AI041810	AI041810 0v34a01.x	C	84	400
18	495	24.3	579	163	BE079600	BE079600 RC5-BT062	C	85	399
19	492	24.2	544	8	AA479605	AA479605 zu42h06.f	C	86	399
20	492	24.2	593	32	AV717337	AV717337 AV717337	C	87	399
21	491	24.0	508	20	AI421428	AI421428 lf25h07.x	C	88	397
22	487	24.0	942	173	BG109874	BG109874 602279542	C	89	395
23	486	23.9	493	16	AI088836	AI088836 q808d08.x	C	90	394
24	480	23.6	480	20	AI419386	AI419386 lf28a05.x	C	91	393
25	480	23.5	480	145	BF196461	BF196461 hr86e07.x	C	92	392
26	478	22.9	478	117	AM166105	AM166105 xf45c09.x	C	93	392
27	465	22.5	473	117	AM513611	AM513611 x047c11.x	C	94	392
28	464	22.8	762	166	BE31617	BE31617 601151459	C	95	391
29	463	22.8	464	111	AM057816	AM057816 wx04c10.x	C	96	391
30	463	22.8	466	7	AA417296	AA417296 zu18c04.s	C	97	388
31	463	22.8	546	122	AM964454	AM964454 EST376527	C	98	386
32	462	22.7	462	17	AI220471	AI220471 9943e03.x	C	99	386
33	456	22.4	462	9	AA576267	AA576267 nm62d08.s	C	100	385
34	452	22.2	452	168	BF733377	BF733377 nae11a03.	C	101	385
35	450	22.1	501	110	AM024459	AM024459 w076f06.x	C	102	383
36	448	22.0	448	110	AM003884	AM003884 ws61f04.x	C	103	383
37	447	22.0	447	19	AI347635	AI347635 q099h05.x	C	104	381
38	447	22.0	451	18	AI304937	AI304937 q020b02.x	C	105	376
39	446	21.9	453	110	AM020149	AM020149 d105a08.y	C	106	375
40	444	21.8	495	12	AA843850	AA843850 a118d06.s	C	107	372
41	443	21.8	460	190	W58242	W58242 zd18c12.s1	C	108	372
42	443	21.8	878	155	BC538574	BC538574 602567303	C	109	372
43	442	21.7	444	11	AA772434	AA772434 a130d06.s	C	110	368
44	441	21.7	441	7	AA470629	AA470629 ne12b06.s	C	111	368
45	440	21.6	521	3	AA164657	AA164657 z098g11.s	C	112	368
46	440	21.6	624	122	AM958900	AM958900 EST370970	C	113	368
47	439	21.6	439	6	AA393599	AA393599 z172e08.r	C	114	363
48	439	21.6	439	7	AA465651	AA465651 aa31a05.s	C	115	363
49	437	21.5	563	148	BF433282	BF433282 nab8904.	C	116	363
50	436	21.4	487	3	AA164656	AA164656 z098g11.r	C	117	363
51	436	21.4	494	114	AM340810	AM340810 h095f02.x	C	118	362
52	432	21.2	433	7	AA437350	AA437350 zw30c10.f	C	119	357
53	432	21.2	433	19	AI344527	AI344527 qp13h12.x	C	120	357
54	432	21.1	457	19	AI333667	AI333667 q908g06.x	C	121	354
55	428	21.1	509	15	AI018757	AI018757 0v78h02.x	C	122	354
56	427	21.0	580	190	W37670	W37670 zc12g03.r1	C	123	352
57	426	21.0	432	4	AA281406	AA281406 z995d10.x	C	124	352
58	425	20.9	425	16	AA159809	AA159809 q073c11.x	C	125	351
59	424	20.9	424	122	AM951698	AM951698 EST373768	C	126	351
60	423	20.8	544	104	AI949241	AI949241 wq20h08.x	C	127	346
61	422	20.8	432	15	AI081927	AI081927 o281e08.x	C	128	346
62	422	20.8	452	117	AM517173	AM517173 x070f09.x	C	129	344
63	422	20.8	503	15	AI031676	AI031676 0w48g05.s	C	130	342
64	421	20.7	421	9	AA554092	AA554092 nk34g02.s	C	131	341
65	420	20.7	424	22	AI573012	AI573012 tm65c05.x	C	132	341
66	418	20.6	423	11	AA770637	AA770637 0e54f01.s	C	133	340
67	417	20.5	417	19	AI366937	AI366937 qz23e02.x	C	134	340
							C	135	339
							C	136	339
							C	137	338
							C	138	337
							C	139	335
							C	140	331
							C	141	331
							C	142	331
							C	143	331
							C	144	331
							C	145	331
							C	146	331
							C	147	331
							C	148	331
							C	149	331
							C	150	331
							C	151	331
							C	152	331
							C	153	331
							C	154	331
							C	155	331
							C	156	331
							C	157	331
							C	158	331
							C	159	331
							C	160	331
							C	161	331
							C	162	331
							C	163	331
							C	164	331
							C	165	331
							C	166	331
							C	167	331
							C	168	331
							C	169	331
							C	170	331
							C	171	331
							C	172	331
							C	173	331
							C	174	331
							C	175	331
							C	176	331
							C	177	331
							C	178	331
							C	179	331
							C	180	331
							C	181	331
							C	182	331
							C	183	331
							C	184	331
							C	185	331
							C	186	331
							C	187	331
							C	188	331
							C	189	331
							C	190	331
							C	191	331
							C	192	331
							C	193	331
							C	194	331
							C	195	331
							C	196	331
							C	197	331
							C	198	331
							C	199	331
							C	200	331
							C	201	331
							C	202	331
							C	203	331
							C	204	331
							C	205	331
							C	206	331
							C	207	331
							C	208	331
							C	209	331
							C	210	331
							C	211	331
							C	212	331
							C	213	331
							C	214	331
							C	215	331
							C	216	331
							C	217	331
							C	218	331
							C	219	331
							C	220	331
							C	221	331
							C	222	331
							C	223	331
							C	224	331
							C	225	331
							C	226	331
							C	227	331
							C	228	331
							C	229	331
							C	230	331
							C	231	331
							C	232	331
							C	233	331
							C	234	331

C 141	329	16.2	329	20	A1439760	A1439760 t161a12.x	214	256	12.6	394	13	AA890442	AA890442 aK10h10.s
C 142	329	16.2	329	103	A1879862	A1879862 ar95603.x	215	256	12.6	439	22	A1571215	A1571215 tK43606.x
C 143	329	16.2	430	159	N27425	N27425 yx81f11.s1	C 216	255	12.5	235	20	A1471226	A1471226 t192d02.x
C 144	329	16.2	423	188	R79812	R79812 y185f06.s1	C 217	255	12.5	445	21	A1493440	A1493440 t91c03.x
C 145	329	16.2	663	33	AV723775	AV723775 AV723775	C 218	254	12.5	513	187	R51314	R51314 yG72e06.s1
C 146	328	16.1	481	23	A1653966	A1653966 tY04a05.x	C 219	251	12.3	251	12	AA788679	AA788679 aH30d03.s
C 147	327	16.1	349	8	AA489232	AA489232 aa57a05.x	C 220	251	12.3	324	115	AAW392267	AAW392267 OVO-ST023
C 148	327	16.1	408	159	N30541	N30541 yx49g04.r1	C 221	251	12.3	375	159	H96444	H96444 yM03e09.r1
C 149	325	16.0	329	14	AA952967	AA952967 oa77a02.s	C 222	251	12.3	480	190	W37530	W37530 zc12g03.s1
C 150	324	15.9	374	8	AA504343	AA504343 oa14d07.x	C 223	251	12.3	779	154	BS529889	BS529889 tS47d03.x
C 151	323	15.9	324	15	A1034047	A1034047 oa14d07.x	C 224	251	12.3	336	27	A1625128	A1625128 tS47d03.x
C 152	322	15.8	439	17	A1242587	AA504343 aa57h03.r	C 225	249	12.2	629	32	AV702380	AV702380 AV702380
C 153	321	15.8	527	3	AA149855	AA149855 zc17e04.s1	C 226	248	12.2	439	32	AV725588	AV725588 AV725588
C 154	319	15.7	366	159	H96859	AA179955 zc17e04.s1	C 227	248	12.2	724	190	W27971	W27971 40D4 Human
C 155	318	15.6	318	17	A1214761	H96859 yw03e09.s1	C 228	245	12.1	724	190	W27971	W27971 40D4 Human
C 156	317	15.6	420	159	N20818	AA1214761 qm30e01.x	C 229	245	12.1	267	15	A1084246	A1084246 yw07f04.s1
C 157	317	15.6	426	19	A1636775	N20818 yx49g04.s1	C 230	244	12.0	468	159	N29628	N29628 yw07f04.s1
C 158	316	15.5	396	190	W15286	A1636775 q941d03.x	C 231	244	12.0	691	155	BS548534	BS548534 602576527
C 159	316	15.5	396	190	W15286	R25336 yH42b03.r1	C 232	243	12.0	260	156	D20253	D20253 HDMG501227
C 160	315	15.5	329	23	A1702447	A1702447 t294d04.x	C 233	243	12.0	260	156	D20253	D20253 HDMG501227
C 161	315	15.5	335	190	M39007	W29007 zB29f03.r1	C 234	242	11.9	242	3	AA169533	AA169533 zc91a04.s
C 162	315	15.5	337	12	AA554083	AA654083 a163d08.s	C 235	242	11.9	338	5	AA298870	AA298870 EST114470
C 163	313	15.4	458	187	R17981	R17981 y916f03.r1	C 236	241	11.9	419	159	N30124	N30124 yx81a06.s1
C 164	312	15.3	400	159	N27239	N27239 yx17d05.r1	C 237	240	11.9	387	190	W42449	W42449 A1349539
C 165	311	15.3	362	159	NS7460	A1963141 wC27h02.x	C 238	240	11.8	269	115	AA638966	AA638966 tS47d03.x
C 166	311	15.3	458	14	AA947478	NS7460 yw88g09.r1	C 239	240	11.8	416	165	BE243164	BE243164 qP73a02.x
C 167	311	15.3	458	14	AA947478	AA947478 oK20h09.s	C 240	239	11.8	540	164	BE169352	BE169352 qP73a02.x
C 168	310	15.2	421	9	AA574012	AA574012 r1f4e08.s	C 241	238	11.7	356	19	A1349539	A1349539 qP73a02.x
C 169	310	15.2	455	188	R62705	R62705 y110d06.r1	C 242	238	11.7	387	190	W42449	W42449 zc22e08.s1
C 170	308	15.2	753	32	AV721338	AV721438 AV721438	C 243	237	11.7	305	13	AA890510	AA890510 aK12d04.s
C 171	306	15.1	306	7	AA416509	AA416509 zV39h06.s	C 244	237	11.7	344	5	AA298831	AA298831 EST114463
C 172	305	15.0	458	7	AA416509	AA636009 n135e09.s	C 245	237	11.7	442	170	BF848965	BF848965 OVI-EM005
C 173	304	15.0	304	155	BS576280	AA636009 zU18c04.r	C 246	237	11.7	442	170	BF848965	BF848965 OVI-EM005
C 174	304	15.0	327	24	A1750027	BS576280 602595783	C 247	237	11.7	442	170	BF848965	BF848965 OVI-EM005
C 175	303	14.9	302	22	A1567231	A1567231 tP25f12.x	C 248	237	11.4	446	32	AV727708	AV727708 t172e03.x
C 176	302	14.9	302	22	A1567231	A1567231 tP25f12.x	C 249	237	11.4	446	32	AV727708	AV727708 t172e03.x
C 177	301	14.8	431	15	AT429304	A1039904 oP52f03.x	C 250	227	11.2	387	157	H12551	H12551 y112b09.s1
C 178	301	14.8	452	189	AT429304	T174223 yC08f106.r1	C 251	227	11.2	407	187	R34849	R34849 yH86c05.s1
C 179	301	14.8	452	189	AT429304	T174223 yC08f106.r1	C 252	227	11.2	407	187	R34849	R34849 yH86c05.s1
C 180	301	14.8	582	159	NA1633	NA1633 yw67f04.r1	C 253	225	11.0	415	142	BE967755	BE967755 601648753
C 181	299	14.7	586	110	AA020210	AA020210 d106a07.y	C 254	223	11.0	423	17	A1244083	A1244083 q163b06.x
C 182	299	14.7	352	13	AA878389	AA878389 o665c09.s	C 255	222	10.9	336	15	A1075460	A1075460 oZ82h11.x
C 183	299	14.7	352	13	AA878389	H10236 yM02a06.r1	C 256	222	10.9	336	15	A1075460	A1075460 oZ82h11.x
C 184	297	14.6	434	159	N35368	N35368 yV23b04.s1	C 257	222	10.9	336	15	A1075460	A1075460 oZ82h11.x
C 185	296	14.6	674	155	BS573861	BS573861 602594993	C 258	222	10.9	336	15	A1075460	A1075460 oZ82h11.x
C 186	295	14.5	295	5	AA298829	AA298829 EST114461	C 259	220	10.8	446	32	AV708922	AV708922 AV708922
C 187	294	14.5	417	159	N24900	N24900 yV01b11.s1	C 260	219	10.8	446	154	BS535906	BS535906 602564071
C 188	291	14.3	477	190	W58241	W58241 zd18c12.r1	C 261	219	10.8	446	154	BS535906	BS535906 602564071
C 189	290	14.3	969	32	AV726680	AV726680 AV726680	C 262	219	10.8	446	154	BS535906	BS535906 602564071
C 190	285	14.0	285	104	A1948455	A1948455 wq06a03.x	C 263	218	10.7	423	13	AA928947	AA928947 oO31e03.s
C 191	284	14.0	318	1	AA025912	AA025912 zc91b07.r	C 264	216	10.6	380	32	AV660751	AV660751 AV660751
C 192	283	13.9	287	159	NA34708	NA34708 yx81f11.r1	C 265	214	10.5	378	187	R44801	R44801 y923f07.s1
C 193	283	13.9	457	187	R51426	R51426 y72e06.r1	C 266	212	10.4	363	188	T33335	T33335 EST57463 Hu
C 194	279	13.7	916	175	BS289998	BS289998 602385149	C 267	209	10.3	843	151	BF665336	BF665336 602118683
C 195	278	13.7	278	19	A1352243	A1352243 q11e11.x	C 268	207	10.2	214	110	AAW020209	AAW020209 d106a06.y
C 196	277	13.6	368	10	AA652023	AA652023 ns47h11.s	C 269	203	10.0	203	102	A1804330	A1804330 tC56a09.x
C 197	275	13.5	338	10	AA652014	AA652014 ns47h11.s	C 270	202	9.9	350	103	AA164333	AA164333 w778d06.x
C 198	274	13.5	386	158	H85240	H85240 yV85d04.r1	C 271	202	9.9	350	103	AA164333	AA164333 w778d06.x
C 199	274	13.5	404	15	A1022495	A1022495 oW50d04.s	C 272	201	9.9	201	3	AA169729	AA169729 zc08b07.s
C 200	271	13.5	445	17	AA459425	AA459425 aa31a05.r	C 273	201	9.9	486	159	NS4515	NS4515 yV37c04.s1
C 201	271	13.5	339	139	BF666506	BF666506 602123958	C 274	200	9.8	245	19	A1349540	A1349540 qP73a04.x
C 202	271	13.3	913	151	BF666506	BF666506 602123958	C 275	199	9.8	159	11	AA782659	AA782659 aJ07c03.s
C 203	270	13.3	419	32	AV725899	AV725899 AV725899	C 276	199	9.8	159	11	AA782659	AA782659 aJ07c03.s
C 204	269	13.2	344	1	AA035073	AA035073 zB27d02.r	C 277	197	9.7	348	144	BF087152	BF087152 PM1-HT045
C 205	269	13.2	344	1	AA035073	AA035073 zB27d02.r	C 278	196	9.6	166	22	A1589478	A1589478 tG76f03.x
C 206	267	13.1	452	173	BG104855	BG104855 602311638	C 279	196	9.6	166	22	A1589478	A1589478 tG76f03.x
C 207	265	13.0	447	159	N76218	N76218 yx37c04.r1	C 280	195	9.6	442	159	NA1531	NA1531 yV04e08.r1
C 208	265	13.0	458	159	N76218	N76218 yx37c04.r1	C 281	194	9.5	336	187	R19136	R19136 y923f07.r1
C 209	264	13.0	315	17	A1205633	A1205633 oX40g10.s	C 282	192	9.4	332	180	T61972	T61972 yV96f03.r1
C 210	261	12.8	261	159	N73409	N73409 EST55e105.B	C 283	190	9.3	344	5	AA298846	AA298846 EST114447
C 211	260	12.7	339	188	R63122	R63122 y101d06.r1	C 284	190	9.3	335	158	H65558	H65558 yV67g05.r1
C 212	258	12.7	441	138	BE620202	BE620202 601483270	C 285	190	9.3	473	190	W52608	W52608 zc49d06.r1
C 213	256	12.6	332	17	A1244926	A1244926 qJ93e12.x	C 286	189	9.3	310	19	A1349546	A1349546 qP73a10.x

287	189	9.3	342	5	AA298119	AA298119	EST113683	360	95	4.7	4.6	322	165	BE265002	BE265002	EST1136832
288	187	9.2	421	188	R79321	R79321	y190c02.s1	361	94	4.6	4.6	463	158	BE6728	BE6728	y08b11.s1
c 289	186	9.1	432	188	R66608	R66608	y132e10.r1	362	92	4.5	4.5	199	169	BE743697	BE743697	PM7-FH072
c 290	185	9.1	538	157	H12550	H12550	y112b09.r1	363	93	4.5	4.5	274	2	AA095751	AA095751	15753_seq
291	186	9.1	185	103	A1880032	A1880032	ap24b06.x	364	87	4.3	4.3	403	102	A1823740	A1823740	wj12b06.x
292	185	9.1	300	5	AA298852	AA298852	EST114452	365	82	4.0	4.0	240	159	H96709	H96709	yc02b09.r1
c 293	185	9.1	433	158	H59133	H59133	yf03f07.s1	366	80	3.9	3.9	246	19	A1381349	A1381349	tc049g09.x
c 300	175	8.6	226	113	A1244187	A1244187	xa52f02.x	367	75	3.7	171	121	AA818193	AA818193	RC1-ST0721	
c 301	174	8.6	218	117	A1244079	A1244079	q163b02.x	372	75	3.7	171	121	AA818193	AA818193	RC1-ST0721	
c 302	174	8.6	218	117	A1244079	A1244079	q163b02.x	373	75	3.7	171	121	AA818193	AA818193	RC1-ST0721	
c 303	174	8.6	218	117	A1244079	A1244079	q163b02.x	376	75	3.7	171	121	AA818193	AA818193	RC1-ST0721	
c 304	173	8.5	281	15	A1051513	A1051513	oxa3d08.s	377	74	3.6	235	174	BC151357	BC151357	na13g605.y	
c 305	172	8.5	957	153	BC428572	BC428572	6024934b7	378	73	3.6	340	223	AA070180	AA070180	HS_3040_A	
c 306	168	8.3	181	2	AA090530	AA090530	y07f12.seq	379	73	3.6	340	223	AA070180	AA070180	HS_3040_A	
c 307	167	8.2	840	150	BE9530606	BE9530606	6020718b1	380	47	2.3	395	10	AA645356	AA645356	ws79g12.r	
c 308	166	8.2	427	187	R24623	R24623	yh36607.r1	381	47	2.3	408	3	AA171248	AA171248	ms43f07.r	
c 309	165	8.1	309	17	A1220862	A1220862	q907f10.x	382	47	2.3	410	113	AA259031	AA259031	um16e09.y	
c 310	163	8.0	303	121	AA868999	AA868999	MR3-SN006	383	47	2.3	435	141	BE851264	BE851264	uw4da03.y	
c 311	163	8.0	303	121	AA868999	AA868999	MR3-SN006	384	47	2.3	435	141	BE851264	BE851264	uw4da03.y	
c 312	159	7.8	207	1	AA035490	AA035490	a12f1f2.s	385	47	2.3	451	168	BE719294	BE719294	ms38f12.s	
c 313	159	7.8	207	1	AA035490	AA035490	a12f1f2.s	386	47	2.3	451	168	BE719294	BE719294	ms38f12.s	
c 314	158	7.8	260	188	R79912	R79912	y185f06.r1	387	47	2.3	468	122	AAW912002	AAW912002	u4f0h04.y	
c 315	158	7.8	260	188	R79912	R79912	y185f06.r1	388	47	2.3	468	122	AAW912002	AAW912002	u4f0h04.y	
c 316	156	7.7	483	158	H56745	H56745	yq9p605.r1	389	47	2.3	469	24	AA839653	AA839653	wq97h03.r	
c 317	155	7.6	321	188	R77591	R77591	y176a11.s1	390	47	2.3	470	4	AA277184	AA277184	uK50a04.y	
c 318	153	7.5	153	11	AA732553	AA732553	n292g09.s	391	47	2.3	481	12	AA815658	AA815658	v47f102.r	
c 319	152	7.5	326	159	NS2199	NS2199	y227f04.s1	392	47	2.3	481	12	AA815658	AA815658	v47f102.r	
c 320	152	7.5	861	168	BF697773	BF697773	602129945	393	47	2.3	497	17	A1226594	A1226594	u100c03.y	
321	146	7.2	260	5	AA298825	AA298825	EST114458	394	47	2.3	497	17	A1226594	AA822963	w39a607.r	
322	146	7.2	260	5	AA298825	AA298825	EST114458	395	47	2.3	497	17	A1226594	AA822963	w39a607.r	
323	146	7.2	260	5	AA298825	AA298825	EST114458	396	47	2.3	497	17	A1226594	AA822963	w39a607.r	
324	143	7.0	208	5	AA297471	AA297471	EST112999	397	47	2.3	502	13	AA867880	AA867880	vx22c09.y	
c 325	142	7.0	844	150	BF570757	BF570757	602075769	398	47	2.3	502	13	AA867880	AA867880	vx22c09.y	
c 326	140	6.9	334	17	A1186656	A1186656	q431h02.x	399	47	2.3	505	140	AA741947	AA741947	ux23f109.x	
c 327	140	6.9	334	17	A1186656	A1186656	q431h02.x	400	47	2.3	505	140	AA741947	AA741947	ux23f109.x	
328	140	6.9	334	17	A1186656	A1186656	q431h02.x	401	47	2.3	505	140	AA741947	AA741947	ux23f109.x	
329	140	6.9	334	17	A1186656	A1186656	q431h02.x	402	47	2.3	508	114	AA318610	AA318610	u02cc09.y	
330	137	6.7	751	155	BE570997	BE570997	602591585	403	47	2.3	517	11	AA755546	AA755546	vs87f029.r	
c 331	135	6.6	380	19	A1372808	A1372808	EST175352	404	47	2.3	517	11	AA755546	AA755546	vs87f029.r	
c 332	132	6.4	333	23	A1689439	A1689439	lx9412.x	405	47	2.3	548	21	A1527733	A1527733	u17a0a08.y	
c 333	130	6.4	332	110	AA023054	AA023054	d148a02.y	406	47	2.3	572	111	AA106795	AA106795	u17a0a08.y	
c 334	130	6.4	340	139	BE077045	BE077045	MRO-HT040	407	47	2.3	577	111	AA106795	AA106795	u17a0a08.y	
c 335	129	6.3	344	32	AAV60547	AAV60547	AAV60547	408	47	2.3	630	107	AU035148	AU035148	u17a0a08.y	
c 336	126	6.2	377	5	A1318832	A1318832	EST12189	409	47	2.3	634	136	BE533490	BE533490	u17a0a08.y	
c 337	125	6.2	841	146	BF214591	BF214591	601845906	410	47	2.3	635	169	BF784590	BF784590	u17a0a08.y	
338	126	6.1	881	145	BF185484	BF185484	601845906	411	47	2.3	635	169	BF784590	BF784590	u17a0a08.y	
339	122	6.0	173	8	AA482519	AA482519	zv05c02.r	412	47	2.3	688	144	BF101430	BF101430	u17a0a08.y	
340	121	6.0	203	159	NS5129	NS5129	y222d12.r1	413	47	2.3	714	149	BF468315	BF468315	u17a0a08.y	
341	118	5.8	478	191	W92614	W92614	ze04a08.r1	414	47	2.3	714	149	BF468315	BF468315	u17a0a08.y	
342	117	5.8	117	13	AA894738	AA894738	oJ28d12.s	415	47	2.3	766	107	AU035395	AU035395	u17a0a08.y	
343	116	5.7	148	191	z20068	z20068	HSAAABTUS.S	416	47	2.3	767	107	AU035395	AU035395	u17a0a08.y	
344	116	5.7	151	191	z20056	z20056	HSAAABTUS.S	417	47	2.3	796	107	AU080227	AU080227	u17a0a08.y	
345	116	5.7	204	191	z20090	z20090	HSAAABTUS.S	418	47	2.3	809	107	AU080111	AU080111	u17a0a08.y	
346	116	5.7	372	157	DB00022	DB00022	H0M004H08A	419	47	2.3	827	107	AU079816	AU079816	u17a0a08.y	
c 347	116	5.7	414	158	H56665	H56665	yq9p605.s1	420	47	2.3	898	175	BC294880	BC294880	u17a0a08.y	
c 348	115	5.7	198	103	A1904513	A1904513	PM-PT057	421	47	2.3	904	175	BC294880	BC294880	u17a0a08.y	
c 349	109	5.4	645	146	BF212134	BF212134	601813557	422	47	2.3	943	174	BC1772247	BC1772247	u17a0a08.y	
c 350	106	5.2	417	2	AA084217	AA084217	zn17h03.r	423	47	2.3	962	174	BC173896	BC173896	u17a0a08.y	
351	103	5.1	105	7	AA419372	AA419372	zv39h06.r	424	47	2.3	968	144	BF119697	BF119697	u17a0a08.y	
352	101	5.0	132	188	T35721	T35721	EST930064.Hu	425	47	2.3	968	144	BF119697	BF119697	u17a0a08.y	
c 353	100	4.9	166	123	AAW94532	AAW94532	RC0-BN003	426	47	2.2	1989	192	AK017482	AK017482	u17a0a08.y	
c 354	100	4.9	169	123	AAW94531	AAW94531	RC0-BN003	427	47	2.2	315	161	BB586499	BB586499	u17a0a08.y	
c 355	100	4.9	274	187	R24520	R24520	yh36d07.s1	428	47	2.2	446	166	BF776911	BF776911	u17a0a08.y	
c 356	99	4.9	758	146	BF246795	BF246795	601855621	429	42	2.1	633	115	AA370869	AA370869	u17a0a08.y	
c 357	98	4.8	338	188	T88945	T88945	yc83f06.s1	430	41	2.0	971	107	AU078791	AU078791	u17a0a08.y	
c 358	96	4.7	229	158	H85873	H85873	yw29a10.s1	431	42	2.1	478	168	BF71854	BF71854	MI-D-E4-	
359	96	4.7	243	158	H85873	H85873	yw29a10.s1	432	38	1.9	122	164	BE172652	BE172652	MRO-HT05	

433	38	1.9	388	113	AW60017	um82d09.y	C 506	21	1.0	252	30	AV418121	AV418121
434	38	1.9	504	116	AW62383	BP230009A	C 507	21	1.0	305	13	AA890510	AA890510
435	38	1.9	513	116	AW64246	BP230009A	C 508	21	1.0	309	39	AA409348	AA409348
436	36	1.8	811	107	AU050986	AU050986	C 509	21	1.0	427	30	AV420110	AV420110
437	35	1.7	243	107	AU059538	AU059538	C 510	21	1.0	429	30	AV426176	AV426176
438	35	1.7	369	115	AW352739	AW352739	C 511	21	1.0	439	30	AV422907	AV422907
439	34	1.7	616	155	BG547804	BG547804	C 512	21	1.0	445	164	BE212330	BE212330
440	33	1.6	331	3	AA200006	AA200006	C 513	21	1.0	461	119	AW179762	AW179762
441	33	1.6	403	11	AA763415	AA763415	C 514	21	1.0	542	230	AO533624	AO533624
442	33	1.6	403	17	AL173768	AL173768	C 515	21	1.0	557	179	BE749749	BE749749
443	33	1.6	406	145	BE199999	BE199999	C 516	21	1.0	570	164	BE212364	BE212364
444	33	1.6	437	122	AW908431	AW908431	C 517	21	1.0	724	170	BF663946	BF663946
445	33	1.6	437	164	BE198768	BE198768	C 518	21	1.0	1101	219	CNS008BT	CNS008BT
446	33	1.6	459	13	AA914720	AA914720	C 519	20	1.0	104	102	AT789946	AT789946
447	33	1.6	472	173	BG101010	BG101010	C 520	20	1.0	119	16	AL137321	AL137321
448	33	1.6	483	163	BE134863	BE134863	C 521	20	1.0	226	31	AV534545	AV534545
449	33	1.6	488	143	BF020773	BF020773	C 522	20	1.0	235	128	BE201952	BE201952
450	33	1.6	495	145	BE194948	BE194948	C 523	20	1.0	236	29	AV337412	AV337412
451	33	1.6	651	166	BE374581	BE374581	C 524	20	1.0	238	156	C92607	C92607
452	33	1.6	922	145	BF166848	BF166848	C 525	20	1.0	245	161	BS800061	BS800061
453	33	1.6	947	169	BF786430	BF786430	C 526	20	1.0	250	135	BA499714	BA499714
454	33	1.6	986	175	BG244389	BG244389	C 527	20	1.0	277	226	AO244652	AO244652
455	33	1.6	1092	175	BG298183	BG298183	C 528	20	1.0	280	161	BS655554	BS655554
456	32	1.6	254	120	AW744313	AW744313	C 529	20	1.0	283	132	BS352081	BS352081
457	32	1.6	292	161	BS578638	BS578638	C 530	20	1.0	287	9	AA585726	AA585726
458	32	1.6	362	141	BE851218	BE851218	C 531	20	1.0	297	128	BR223259	BR223259
459	32	1.6	397	162	BE002340	BE002340	C 532	20	1.0	300	31	AV561170	AV561170
460	32	1.6	424	115	AW413298	AW413298	C 533	20	1.0	302	161	BS559820	BS559820
461	32	1.6	504	103	AI930100	AI930100	C 534	20	1.0	305	119	AM659824	AM659824
462	32	1.6	539	119	AW702048	AW702048	C 535	20	1.0	320	128	BR209124	BR209124
463	32	1.6	582	166	BE309388	BE309388	C 536	20	1.0	321	150	BF554768	BF554768
464	32	1.6	661	165	BE291891	BE291891	C 537	20	1.0	333	151	BS544768	BS544768
465	32	1.6	696	169	BF788565	BF788565	C 538	20	1.0	335	187	RS4442	RS4442
466	32	1.6	866	137	BE569816	BE569816	C 539	20	1.0	351	114	AW307748	AW307748
467	32	1.6	887	145	BF178365	BF178365	C 540	20	1.0	359	21	AI553421	AI553421
468	32	1.6	889	137	BE573270	BE573270	C 541	20	1.0	370	107	AU061974	AU061974
469	32	1.6	909	144	BF100402	BF100402	C 542	20	1.0	426	15	AI036061	AI036061
470	32	1.6	928	145	BF143123	BF143123	C 543	20	1.0	435	31	AV563809	AV563809
471	32	1.6	932	143	BF030300	BF030300	C 544	20	1.0	444	225	AO220484	AO220484
472	31	1.5	454	13	AA882096	AA882096	C 545	20	1.0	456	120	AM757376	AM757376
473	31	1.5	480	123	AA985491	AA985491	C 546	20	1.0	456	153	BG450030	BG450030
474	31	1.5	888	144	BF100333	BF100333	C 547	20	1.0	482	30	AV440673	AV440673
475	30	1.5	410	168	BF711459	BF711459	C 548	20	1.0	491	31	AV564781	AV564781
476	30	1.5	789	137	BE573048	BE573048	C 549	20	1.0	497	235	AO882573	AO882573
477	29	1.4	139	2	AA072602	AA072602	C 550	20	1.0	519	297	AZ017722	AZ017722
478	29	1.4	370	10	AA647140	AA647140	C 551	20	1.0	536	249	AZ774776	AZ774776
479	29	1.4	965	144	BF123569	BF123569	C 552	20	1.0	537	168	BF703203	BF703203
480	28	1.4	101	172	BF987085	BF987085	C 553	20	1.0	539	14	AA275209	AA275209
481	28	1.4	265	120	AW784274	AW784274	C 554	20	1.0	559	31	AV539155	AV539155
482	27	1.3	158	102	AI800370	AI800370	C 555	20	1.0	565	144	BR079749	BR079749
483	27	1.3	571	120	AM766059	AM766059	C 556	20	1.0	597	236	B29238	B29238
484	26	1.3	127	151	BF660654	BF660654	C 557	20	1.0	605	31	AV559166	AV559166
485	26	1.3	312	128	BR205593	BR205593	C 558	20	1.0	609	13	AA892793	AA892793
486	26	1.3	450	12	AA769596	AA769596	C 559	20	1.0	651	230	AO540871	AO540871
487	26	1.3	467	141	BE861511	BE861511	C 560	20	1.0	666	167	BE440606	BE440606
488	25	1.2	381	145	BF190252	BF190252	C 561	20	1.0	673	20	AI451342	AI451342
489	24	1.2	216	146	BF282589	BF282589	C 562	20	1.0	706	165	BE274533	BE274533
490	24	1.2	239	17	AL175972	AL175972	C 563	20	1.0	711	246	AZ611078	AZ611078
491	24	1.2	289	16	AL117327	AL117327	C 564	20	1.0	732	121	BE977080	BE977080
492	24	1.2	385	14	AA997135	AA997135	C 565	20	1.0	780	152	BS6596487	BS6596487
493	24	1.2	394	15	AI050315	AI050315	C 566	20	1.0	811	239	AZ198039	AZ198039
494	24	1.2	479	19	AI407881	AI407881	C 567	20	1.0	816	227	AO380124	AO380124
495	24	1.2	620	14	AI012226	AI012226	C 568	20	1.0	862	220	CNS010SB	CNS010SB
496	24	1.2	682	122	AW916468	AW916468	C 569	20	1.0	885	245	AZ533320	AZ533320
497	24	1.2	744	114	AM335297	AM335297	C 570	20	1.0	885	245	AZ533396	AZ533396
498	23	1.1	149	244	AZ487189	AZ487189	C 571	20	1.0	989	227	AZ678330	AZ678330
499	23	1.1	270	137	BE176694	BE176694	C 572	20	1.0	1118	241	CNS031L7	CNS031L7
500	22	1.1	203	139	BE571947	BE571947	C 573	20	1.0	1926	192	AK003898	AK003898
501	22	1.1	508	102	AT786626	AT786626	C 574	20	1.0	2310	192	AK009375	AK009375
502	22	1.1	517	138	BE664648	BE664648	C 575	20	1.0	132	145	BF148394	BF148394
503	22	1.1	591	11	AI232614	AI232614	C 576	20	1.0	182	175	BS306344	BS306344
504	22	1.1	792	145	BF142579	BF142579	C 577	20	1.0	200	115	AM429655	AM429655
505	21	1.0	143	168	BF711680	BF711680	C 578	19	0.9	208	117	AM509314	AM509314

[illegible]

C 725	18	0.9	259	131	BB344125	798	18	0.9	362	169	BF763667
C 726	18	0.9	260	109	AV038295	799	18	0.9	362	32	AV656705
C 727	18	0.9	262	128	BB221732	800	18	0.9	363	121	AM833872
C 728	18	0.9	265	169	BF761102	801	18	0.9	363	187	RI10910
C 729	18	0.9	266	14	AA961949	802	18	0.9	364	110	AV749719
C 730	18	0.9	266	21	AI91616	803	18	0.9	366	169	BF761386
C 731	18	0.9	266	169	BF761372	804	18	0.9	366	256	BI7207
C 732	18	0.9	268	257	B92226	805	18	0.9	367	165	BE272851
C 733	18	0.9	269	5	AA310893	806	18	0.9	367	103	AI910646
C 734	18	0.9	269	6	AA373137	807	18	0.9	369	112	AM193293
C 735	18	0.9	270	168	RR2472	808	18	0.9	369	112	AM193293
C 736	18	0.9	274	123	BB000769	809	18	0.9	370	187	R33728
C 737	18	0.9	275	154	BG528687	810	18	0.9	371	143	BF004147
C 738	18	0.9	276	161	BB573914	811	18	0.9	372	32	AV659111
C 739	18	0.9	277	131	BB344917	812	18	0.9	374	14	AI013222
C 740	18	0.9	277	131	BB344918	813	18	0.9	375	105	AL1372838
C 741	18	0.9	282	25	AV145924	814	18	0.9	375	121	AM833921
C 742	18	0.9	284	25	AV137544	815	18	0.9	375	156	C16220
C 743	18	0.9	284	131	BB347220	816	18	0.9	375	162	BE061166
C 744	18	0.9	286	158	HE9537	817	18	0.9	377	159	BE59089
C 745	18	0.9	287	133	BB430052	818	18	0.9	377	244	A2450189
C 746	18	0.9	288	241	A2301973	819	18	0.9	380	172	BG004682
C 747	18	0.9	290	160	BB517516	820	18	0.9	381	151	BF646944
C 748	18	0.9	292	157	DD9418	821	18	0.9	382	16	AI093782
C 749	18	0.9	292	170	BF848536	822	18	0.9	382	226	AO293843
C 750	18	0.9	293	22	AI619400	823	18	0.9	384	238	A2125113
C 751	18	0.9	295	245	AA560482	824	18	0.9	386	1	AA053608
C 752	18	0.9	296	133	BB426395	825	18	0.9	386	113	AM208047
C 753	18	0.9	297	1	AA011422	826	18	0.9	386	140	BE800531
C 754	18	0.9	297	134	BB454082	827	18	0.9	386	172	BE991068
C 755	18	0.9	298	140	BE805128	828	18	0.9	387	256	A2932105
C 756	18	0.9	300	132	BB360271	829	18	0.9	388	224	AO130392
C 757	18	0.9	301	128	BB228946	830	18	0.9	390	24	AI749985
C 758	18	0.9	301	157	DD2745	831	18	0.9	390	32	AV684139
C 759	18	0.9	304	146	BF229852	832	18	0.9	393	2	AA085841
C 760	18	0.9	304	162	BB605530	833	18	0.9	393	104	AI975444
C 761	18	0.9	305	157	DD9427	834	18	0.9	393	159	NE7091
C 762	18	0.9	307	130	BB308451	835	18	0.9	393	128	AO455634
C 763	18	0.9	308	28	AA929657	836	18	0.9	396	111	AM695323
C 764	18	0.9	309	142	BE938956	837	18	0.9	398	14	AI003196
C 765	18	0.9	309	142	BE938973	838	18	0.9	398	191	WB9069
C 766	18	0.9	311	157	DD2728	839	18	0.9	399	148	BF439881
C 767	18	0.9	311	157	DD2744	840	18	0.9	399	162	BE061774
C 768	18	0.9	314	171	BF904849	841	18	0.9	399	240	A2252561
C 769	18	0.9	320	127	BB172741	842	18	0.9	400	10	AA634994
C 770	18	0.9	322	169	BF801048	843	18	0.9	400	108	AV016206
C 771	18	0.9	326	157	DD1751	844	18	0.9	401	10	AA693952
C 772	18	0.9	327	256	B39440	845	18	0.9	402	187	RI10126
C 773	18	0.9	328	251	AA923867	846	18	0.9	403	122	AM894781
C 774	18	0.9	330	118	AM602839	847	18	0.9	403	223	AO013448
C 775	18	0.9	330	157	DD9461	848	18	0.9	404	32	AV645994
C 776	18	0.9	331	136	BE505809	849	18	0.9	405	104	AI974507
C 777	18	0.9	332	111	AM103286	850	18	0.9	405	170	BE822981
C 778	18	0.9	336	174	BG013082	851	18	0.9	405	256	B53857
C 779	18	0.9	337	134	BB453600	852	18	0.9	407	170	BF885791
C 780	18	0.9	339	121	AA821112	853	18	0.9	409	2	AA128504
C 781	18	0.9	339	151	BF635785	854	18	0.9	409	10	AA128504
C 782	18	0.9	342	170	BF885789	855	18	0.9	410	121	AM833919
C 783	18	0.9	343	8	AA480392	856	18	0.9	410	121	AM833919
C 784	18	0.9	343	248	AA721855	857	18	0.9	410	219	AC021285
C 785	18	0.9	344	248	AA721873	858	18	0.9	412	113	AM241283
C 786	18	0.9	345	169	BF761385	859	18	0.9	413	32	AV646303
C 787	18	0.9	346	14	AA972119	860	18	0.9	415	233	AO795452
C 788	18	0.9	347	118	AA618625	861	18	0.9	415	31	AV540369
C 789	18	0.9	347	224	AA083489	862	18	0.9	416	20	AI1443136
C 790	18	0.9	350	188	R93317	863	18	0.9	416	30	AV388503
C 791	18	0.9	353	118	AA630365	864	18	0.9	416	121	AM833881
C 792	18	0.9	354	234	AA812015	865	18	0.9	416	150	BE566094
C 793	18	0.9	357	7	AA429189	866	18	0.9	417	11	AA740980
C 794	18	0.9	357	22	AA1576008	867	18	0.9	417	112	AA9910168
C 795	18	0.9	359	11	AA766111	868	18	0.9	419	119	AA694559
C 796	18	0.9	359	147	BF369675	869	18	0.9	419	158	H53683
C 797	18	0.9	360	170	BF869692	870	18	0.9	419	189	T70337

BF763667	CW0-CS004
AV656705	AV656705
AM833872	QVO-TT000
RI0910	YF38903.s1
AV749719	AV749719
BF761386	RC2-CS001
B17207	344A22.TPB
AI026027	OW03F02.s
BE272851	601171094
AI910646	LT25N02.x
AM193293	X173N05.x
R33728	YH78N08.r1
BF004147	ESY432645
AV659111	AV659111
AI013222	EST207897
AL372838	MEBA53H10
AM833921	QVO-TT000
C16220	C16220
BE061166	QVO-BT004
ME59089	YF57C03.s1
A2450189	IM0246L17
BG004682	MR3-GN018
BF646944	NF076E08E
AI093782	qaz3b03.s
AO293843	HS-2186.B
A2125113	OS2NBD007
AA053608	Z173608.r
AM208047	MI11078e
BE800531	sg94101.Y
BF991068	CM1-GN016
A2932105	474.dh294
AO130392	HS-3017.B
AI749985	at34g06.x
AV684139	AV684139
AA085841	zr63f04.r
AI975444	EST270038
NE7091	YF45T06.s1
AO455634	HS-5156.B
AM695323	NF094A02S
AI003196	am12a06.s
WB9069	zh71H02.s1
BF439881	nab71c08.
BE061774	RC1-BT025
A2252561	RPTC-23-7
AA634994	ab47b12.r
AV016206	AV016206
AA693952	z153d03.s
RI10126	YF31B06.s1
AM894781	CM1-NN003
AO013448	RPTC11-24
AV645994	AV645994
AI974507	T110457e
BF822981	MRI-RT007
B53857	CLT-HSR-201
BF885791	PM4-TN011
AA128504	zm24908.r
AA700745	z162C03.s
AM833919	QVO-TT000
AM833952	QVO-TT000
AG021285	Oryza sat
AM241283	xa54c11.x
AV646303	AV646303
AO795452	nxbD0056M
AV540369	AV540369
AI1443136	sa84g08.Y
AV388503	SA385503
AM833881	QVO-TT000
BE566094	UT-R-BT1-
AA740980	n205e03.s
AA9910168	ut75T03.Y
AA694559	NF077603S
H53683	YU38F07.r1
T70337	Yd13f12.s1



871	18	0.9	420	121	AM833879	QV0-TT000	944	18	0.9	466	168	BP725865	BF725865 bx21a03.y
872	18	0.9	421	6	AA394177	z456a06.r	945	18	0.9	466	187	R54888	R54888 yj74907.r1
873	18	0.9	421	141	BP884750	601506768	946	18	0.9	466	250	AA810133	AA810133 2M0074E02
874	18	0.9	422	224	AO109741	CIT-HSP-2	947	18	0.9	468	22	AI582237	AI582237 tq65c09.x
875	18	0.9	422	229	AO486760	RPCT-11-2	948	18	0.9	468	117	AM527728	AM527728 UT-R-BT1-
876	18	0.9	424	2	AA308277	EST1791177	949	18	0.9	469	20	AI467153	AI467153 vd25c09.x
877	18	0.9	425	22	AI620661	tub5401.x	950	18	0.9	470	22	AI591218	AI591218 tw13b06.x
878	18	0.9	425	113	AM208099	M11198e	951	18	0.9	472	147	BP324384	BP324384 CM2-FT012
879	18	0.9	425	190	W85685	zh55911.s1	952	18	0.9	472	164	BE205878	BE205878 sb48a12.y
880	18	0.9	426	116	AM446926	877877.MAR	953	18	0.9	472	164	BE205878	BE205878 db48d06.x
881	18	0.9	426	158	H75840	yub60410.s1	954	18	0.9	472	147	BP324384	BP324384 CM2-FT012
882	18	0.9	427	110	AM033147	AM033147	955	18	0.9	472	147	BP324384	BP324384 CM2-FT012
883	18	0.9	427	121	AM833953	QV0-TT000	956	18	0.9	473	3	AA179590	AA179590 zpd9611.x
884	18	0.9	427	240	AA2358620	RPCT-23-1	957	18	0.9	473	16	AI094854	AI094854 ga09a02.x
885	18	0.9	428	116	AM451392	UT-H-B13-	958	18	0.9	473	117	AM565289	AM565289 LGL 334.C
886	18	0.9	428	167	BE461162	EST412581	959	18	0.9	473	222	FR0026674	FR0026674
887	18	0.9	428	241	AZ307039	AL3707039	960	18	0.9	474	238	AZ097336	AZ097336
888	18	0.9	429	105	AL371886	MTBA47B04	961	18	0.9	478	105	AL371716	AL371716 MTBA46A07
889	18	0.9	429	258	FR004778	Fugu tubt	962	18	0.9	478	158	H57863	H57863
890	18	0.9	430	113	AM242210	AM242210	963	18	0.9	478	162	BE006765	BE006765 RCS-BN013
891	18	0.9	430	235	AO933811	AO933811	964	18	0.9	479	106	AL563892	AL563892
892	18	0.9	431	172	BE009716	QV1-GN031	965	18	0.9	479	167	BE453743	BE453743
893	18	0.9	431	229	AO459000	HS-5068_B	966	18	0.9	480	2	AA100690	AA100690
894	18	0.9	432	158	H25719	Y154C10.r1	967	18	0.9	480	114	AM281970	AM281970
895	18	0.9	432	169	BF818795	OV2-C1010	968	18	0.9	480	245	AZ515476	AZ515476
896	18	0.9	433	15	AI038031	AT038031	969	18	0.9	481	1	AA002175	AA002175
897	18	0.9	433	114	AM305298	XV93B02.x	970	18	0.9	481	11	AA779093	AA779093
898	18	0.9	434	232	AO686637	nbx0072F	971	18	0.9	481	159	N70812	N70812
899	18	0.9	436	138	H25708	y154a11.r1	972	18	0.9	482	110	AM008401	AM008401
900	18	0.9	436	190	W31249	zbd45c01.r1	973	18	0.9	482	239	AZ162604	AZ162604
901	18	0.9	437	162	BE016611	FK66a11.y	974	18	0.9	483	148	BF425601	BF425601
902	18	0.9	438	189	W04970	z442c03.r1	975	18	0.9	484	112	AM150307	AM150307
903	18	0.9	439	239	AZ199123	SP.1039_A	976	18	0.9	485	150	BP518683	BP518683
904	18	0.9	441	6	AA373526	EST85687	977	18	0.9	487	234	AO856945	AO856945
905	18	0.9	441	20	AA45579	z126a05.x	978	18	0.9	488	111	AM072637	AM072637
906	18	0.9	442	175	BG306754	fm06f06.y	979	18	0.9	489	189	T63023	T63023
907	18	0.9	442	1	AA070722	zmb6f08.s	980	18	0.9	490	112	AM135176	AM135176
908	18	0.9	444	8	AA542672	fa06d01.r	981	18	0.9	490	149	BF514172	BF514172
909	18	0.9	444	228	AO431815	SP.0044_B	982	18	0.9	491	105	AL374360	AL374360
910	18	0.9	444	239	AZ145173	SP.0044_A	983	18	0.9	491	230	AO562643	AO562643
911	18	0.9	445	112	AM170648	x179g10.x	984	18	0.9	491	247	AZ658901	AZ658901
912	18	0.9	445	156	C84417	CR4417	985	18	0.9	492	148	BF439343	BF439343
913	18	0.9	447	5	AA320866	EST23353	986	18	0.9	493	2	B32508	B32508
914	18	0.9	447	15	AI052599	oz10a06.x	987	18	0.9	493	2	AA086003	AA086003
915	18	0.9	448	223	AO051802	RPCT11-52	988	18	0.9	493	110	AM010093	AM010093
916	18	0.9	450	188	AO430493	HS-5061_B	989	18	0.9	493	164	BE206285	BE206285
917	18	0.9	451	188	R63634	y109e12.r1	990	18	0.9	494	224	AO120533	AO120533
918	18	0.9	451	223	AO052469	RPCT11-52	991	18	0.9	494	228	AO415544	AO415544
919	18	0.9	452	188	R77861	y177e03.r1	992	18	0.9	495	137	BE579175	BE579175
920	18	0.9	452	12	AA828424	oc46a12.s	993	18	0.9	495	188	R84749	R84749
921	18	0.9	453	110	AM022892	d145f102.y	994	18	0.9	496	152	BG339893	BG339893
922	18	0.9	454	142	BE947539	UT-M-BH3-	995	18	0.9	497	30	AV392936	AV392936
923	18	0.9	454	234	AO813576	HS-5315_A	996	18	0.9	497	139	BE709856	BE709856
924	18	0.9	454	243	AZ438114	1M0226G20	997	18	0.9	497	151	BF639204	BF639204
925	18	0.9	456	17	AI248611	qh77a09.x	998	18	0.9	498	224	AO148642	AO148642
926	18	0.9	456	230	AO558895	HS.2089_B	999	18	0.9	498	238	AZ125995	AZ125995
927	18	0.9	458	8	AA487510	aa95c11.s	1000	18	0.9	498	242	AZ350641	AZ350641
928	18	0.9	458	110	AM02649	wub6f093.x							
929	18	0.9	459	156	C90966	D1C1							
930	18	0.9	461	1	AA071014	zmb6f08.r							
931	18	0.9	461	246	AZ620677	1M0435E05							
932	18	0.9	462	105	AL500824	AL500824							
933	18	0.9	462	112	AM164836	se78e09.y							
934	18	0.9	462	246	AZ612811	1M0439F22							
935	18	0.9	463	19	AI364327	qy63g05.x							
936	18	0.9	463	116	AM438880	x105c08.x							
937	18	0.9	464	2	AA135221	z627n07.s							
938	18	0.9	464	112	AM166185	x146g01.x							
939	18	0.9	464	229	AO484996	HS.-5191_A							
940	18	0.9	464	230	AO560390	HS.-2077_B							
941	18	0.9	464	233	AO809049	HS.-5443_C							
942	18	0.9	466	122	AM922012	EST353316							
943	18	0.9	466	122	AM922012	EST353316							

## ALIGNMENTS

## RESULT 1

## LOCUS

## DEFINITION

## ACCESSION

6025591691 NIH\_MGC\_61 Homo sapiens cDNA clone IMAGE:4697177 5',  
 mRNA sequence.  
 BGS30418  
 BGS30418.1 GI:13521955

## KEYWORDS

## SOURCE

human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 681)  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 NIH-MGC http://mgi.nci.nih.gov/  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL Contact: Robert Strausberg, Ph.D.  
 COMMENT Email: cgaabs-remail.nih.gov

Tissue Procurement: ATCC  
 CDNA Library Preparation: CLONTECH Laboratories, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L1CM1525 row: i column: 18  
 High quality sequence stop: 681.

## FEATURES

Location/Qualifiers

1..681  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4697177"  
 /clone\_1ib="NIH-MGC\_61"  
 /tissue\_type="embryonal carcinoma"  
 /lab\_host="DH10B (TI phage-resistant)"  
 /note="Organ: testis; Vector: pCMV-LIB (Clontech); Site\_1: SfiI (ggccgctggcc); Site\_2: SfiI (ggccatagcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCGCATATGCGC-3' and 3' adaptor sequence: 5'-ATCTAGAGCGAGCGCGCATAG-3' (30)NN-3' (where B = A, C, G and N = A, C, G, or T). Average insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."

BASE COUNT 176 a 149 c 157 g 199 t

## ORIGIN

Query Match 32.1%; Score 652; DB 154; Length 681;  
 Best Local Similarity 100.0%; Pred. NO. 0;  
 Matches 652; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

338 agggagtgacaacaactgctctacccccagcagtgctcattcgtacccttgaggaatg 397  
 |||||||  
 24 AGGAGTGAACAACTGGCTTACCCCGAGAGTGTATTCGTACCCCTTGAGGATG 83  
 |||||||  
 398 cagttccttcttgcttgcagtggttgcaaatcattcattccttatttctggagaa 457  
 |||||||  
 84 CAGTCTCTTTAGTCTTGCAGTGTGCAATTCATTCACCTCTTATTTCTGAGGAAA 143  
 |||||||  
 458 ctccgttcttgcttgcagtggttgcagtgagaaagtgatattgtaggaagaa 517  
 |||||||  
 144 CTCTGTGTTGGAGTGGCTCCAGTGAAGAGTGTATATGAGGAGGAGGAA 203  
 |||||||  
 518 actaagtggttgaagaccttcagtaaccttgcccaactccgtaactcgccttgcaag 577  
 |||||||  
 204 ACTGAGTGTGAACACCTTTCACTACCTTGGCCACTCTCGTAATCCGCTTTGAG 263  
 |||||||  
 578 aaaactgtctcagttcactcccccaattctctgagtagaagaatgaagtgcac 637  
 |||||||  
 264 AAAACTGTGTCAGTTCACCTCCCTCAATTCCTGAGTGAACAAATGAAGTTGACC 323  
 |||||||  
 638 tgccttcttcttcaactgaagtgaagtgaatgaatgaatgaatgaatgaatgaat 697  
 |||||||  
 324 TGCTCTTTCTTTCTGACTGCAAGTGTACATGATATTTCAAGTTGCTGCTCATTA 383  
 |||||||  
 698 agcatctagcaagaatcattctcgtattattattcactcagtaggaagtgaggttg 757  
 |||||||  
 384 AGCATTTAGCAAGATCATCTCTGATTTATATTCACTGAGACGTGCGAGGTTGATG 443  
 |||||||  
 758 aaatgtggaagcggtatagggaagcctcgaacaattcagtagatgcttcttaagatc 817  
 |||||||

DB 444 AAATTGGAGACGTTATGAGGAGAACTGTGAACATTCAGAGATGCTTCTAAGTCTTG 503  
 |||||||  
 QY 818 ttgaagctctgcaaaagtttgcaagatgacatgacatgacatgacatgacatgac 877  
 |||||||  
 DB 504 TTGACCTCTGCAAAAGTTTGCAGATGACATGTACATCTTTATGTGGAAATGCGTGG 563  
 |||||||  
 QY 878 taagattgactcgtcaagatgacatgacatgacatgacatgacatgacatgacatgac 937  
 |||||||  
 DB 564 TAAGTGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 623  
 |||||||  
 QY 938 ttgaagcaacaagcgaagaagcgaagaagcgaagaagcgaagaagcgaagaagcga 989  
 |||||||  
 DB 624 TTGAGCAAAACAAAGCGAAGACCGACGAGTCTCTTATACCTTCATTTAA 675  
 |||||||

## RESULT 2

LOCUS BG288867 665 bp mRNA EST 21-FEB-2001  
 DEFINITION BG288867.1 GI:13044136  
 mRNA sequence.  
 ACCESSION BG288867  
 VERSION BG288867.1  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 665)  
 NIH-MGC http://mgi.nci.nih.gov/  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL Contact: Robert Strausberg, Ph.D.  
 COMMENT Email: cgaabs-remail.nih.gov

Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L1AM10398 row: p column: 02  
 High quality sequence stop: 665.

## FEATURES

Location/Qualifiers

1..665  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4513009"  
 /clone\_1ib="NIH-MGC\_93"  
 /tissue\_type="transitional cell papilloma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: bladder; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH-MGC Library."

BASE COUNT 185 a 128 c 145 g 207 t

## ORIGIN

Query Match 31.6%; Score 642; DB 175; Length 665;  
 Best Local Similarity 100.0%; Pred. NO. 0;  
 Matches 642; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

556 ctccgtaacgcgctgttcaagaacactgttctcagttcattcccccctcaattctctg 615  
 |||||||  
 DB 1 CTCGTGATAGGCTCTTTCAAGAAACCTGTGTCAGTTCACCTCCCTCAATTCCTG 60  
 |||||||  
 QY 616 agtagaacaatgaagttgacgtgctcttcttcttgaactgaagtgtcacaagatatt 675  
 |||||||  
 DB 61 AGTAGAACAATGAAGTGTGACCTGCTCTTTCTTCTGAACTGCAAGTGTACATGATATT 120  
 |||||||  
 QY 676 tcaagcttgctctcgtcaatgaacatctagcaagaagatcattctcgtattattatca 735  
 |||||||

```

|||||
Db 121 TCAAGCTGCTGCTGCTCAACGATCTTAGCCAGGATCTTCTCTATTTATTTCA 180
Qy 726 ctgagctgagcaggttggatgaatgggaagcgtatgggaagactctgaacattc 795
Db 181 CTGAGAGCTGCGAGGTTTGATGAAATGGGAGCGCTTATGGGAGACTCTGAACAAATTC 240
Qy 796 agaatgctctcaaatccttgtagcgtctgcaaaagtctgagatgacatgacagt 855
Db 241 AGAATGCTCTTAAATCTCTGTTGAGCGCTGCAAAAGTTTGCAATGACATGACAGT 300
Qy 856 cttaatggtgggaatgagcagtgtagagtagtgcacgtcgaatcattgacacccctc 915
Db 301 CTTTATGCTGAGGAAATGAGTGTAGATGATGATGATGATGATGATGATGATGATGAT 360
Qy 916 attgagaagacaaggaactatccttgagggcaaaagaggaagaccacgaagtcctcat 975
Db 361 ATTGAGAGCAAGGACTATCTTGAAGCAAAACAGGAGAACCCAGCAAGTCCCTAT 420
Qy 976 aaacttgacataagataatttgaataatccgtggttctcaacatgtagtctgata 1035
Db 421 AACCTTGCAATATGATATATTTGAAATTCCTGCTTCAACATGATGATGATGATGAT 480
Qy 1036 atgagcgccttgagccttgagcgtgagatcaccctcttcaaatatttggagacatgag 1095
Db 481 ATGATGCGCTTGCGCTTGCTGCTGATATACCTCTTCAATATTTGGAACATGATGAT 540
Qy 1096 ggaatgataatgacatattatagatgacaaacaggaatctgaaatgtagtga 1155
Db 541 GGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
Qy 1156 cctgagcagaattgaaagaggggttggaaatgagcgtt 1197
Db 601 CCTGTGCGAATTTGAAAGAGGGGCTTGAATTTGCTGCTT 642

RESULT 3
LOCUS BG501143 852 bp mRNA EST 27-MAR-2001
DEFINITION 602546382F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4668621 5',
mRNA sequence.
ACCESSION BG501143
VERSION BG501143.1 GI:13462660
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 852)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rgs@bbs.fda.nih.gov
Tissue Procurement: DCTD/DPV
CDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM1479 row: c column: 22
High quality sequence stop: 781.
Location/Qualifiers
1. 852
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4668621"
/clone_1b="NIH_MGC_60"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: prostate; Vector: pDNR-LIB (Clontech);
Site_1: SfiI (ggccgctggcc); Site_2: SfiI (ggccattatggc

```

```

); Double-stranded cDNA was prepared from cell line RNA,
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCCATTTATGACC-3' and 3' adaptor
sequence: 5'-ATCTGAGGCGCCGAGCGCCGAGCATG-dT(30)BN-3'
(where B = A, C, G, or T). Average
insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH-MGC
Library."
BASE COUNT 256 a 131 c 183 g 282 t
ORIGIN
Query Match 30.0%; Score 609; DB 154; Length 852;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 648 ttctgaactgcaagtcacatgatalttcaagctgctgctgcataaagcatcagc 707
Db 1 TTCTGAACCTGCAAGTGTACATGATATTTCAGCTTGTGCTGCTGCTGCTGCTGCTGCT 60
Qy 708 caaagatcattctcctgattatattcactgagctgagcaggtttgagaaattggaa 767
Db 61 CAAAGATCATCTCTGATTTATTTACCTGAGGCTGGAGGTTGGATGAATTTGGAA 120
Qy 768 gctgtaggggaagactctgacaattcagagatgcttctaaagatcctgttgaagcct 827
Db 121 GCGTTATGGGGAAGACTCTGCAACATTCAGAGATCTTTAAGATCTTGTGAGCGCT 180
Qy 828 gcaaaagtctgagatgacatgacatgctttagtgggaagcagtgtagagtagt 887
Db 181 GCAAAATTTGCAAGTATGACATGATGATGATGATGATGATGATGATGATGATGATGAT 240
Qy 888 cactgtaagatcatttgacacctcctcattagagaagaagacatccttgagga 947
Db 241 CACTGTCAAGTATTTGACACCTCCCTCATTTAGGAAGCAAGGATATCTTGAAGCAAA 300
Qy 948 acaagcgaagacccaaggaagtcctcattacactgcatgataaatttgaatattc 1007
Db 301 ACAAGCGAAGAACCCAGCAAGTCCCTATTAACCTTGCAATGATGATGATGATGATGAT 360
Qy 1008 cgtgatttccaacatgacttggatgaatgacgccttgagccttgagcgtgtagtattac 1067
Db 361 CGTGTGTTTCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
Qy 1068 ctcttcaaatatttggaacatgagatccttgatgataatgataatgataatgataat 1127
Db 421 CTCTTCAATATTTGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
Qy 1128 ccaagagattcgaatgagatgagatgacatgacatgacatgacatgacatgacatgac 1187
Db 481 CCAAGAGATTTGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
Qy 1188 ttgagcttttgaatattatccttgatgagcttgaatgagcttgaatgagcttgaatgag 1247
Db 541 TTGCTGTTTGTAAATATATCTTTAGTGTCTTTAAAGTATGATGATGATGATGATGAT 600
Qy 1248 ttataaag 1256
Db 601 TTTATTAAG 609

RESULT 4
LOCUS BE897106 947 bp mRNA EST 20-OCT-2000
DEFINITION 6014396363F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3924687 5',
mRNA sequence.
ACCESSION BE897106
VERSION BE897106.1 GI:10362234
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 947)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC/DCTD/DRP  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LAM9762 row: n column: 16  
High quality sequence stop: 656.

## FEATURES

Source  
Location/Qualifiers  
1..947  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3924687"  
/clone\_lib="NIH-MGC.72"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pCMV-SPORE6; Site: 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life  
Technologies."

BASE COUNT 274 a 155 c 210 g 307 t 1 others  
ORIGIN

Query Match 29.6%; Score 602; DB 141; Length 947;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 602; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

580 aactcgtctcagtcaccccccaccccaatctctcgtagtaggaacaatgaagctgacg 639  
1 AACCTGCTCTCAGTTCACCTCCCTCAATCTCTGAGTAGAACAATGAACTGACCTG 60  
640 cttctctcttcgaaactgaagtgctacagatgatttgaagcttgcctgctgataag 659  
61 CTCTCTCTTCTTGCAGTCAAGTGTCTACATGATTTTAAAGCTTCTGCTGCTATAG 120  
700 catctagcaagatcattctcctgattatatactcagctgagctgaggttggatgaa 759  
121 CATCTACCAAGGATCATCTCTGATTTATATTCACGAGCTGCGAGTTTGGATGAA 180  
760 attggaagcgttatagggaagactcgaacaattcagaagatgcttctaagactcgtt 819  
181 ATTGGAGCGTTATGGGAGAGACTCTGAACAATTCAGAGATGCTTCTTAAGATCCTTGT 240  
820 gaagctctgcaaaagttgcagatgacatgacagctcttattggtggaaagcagtgta 879  
241 GACGCTCTGGAAGATTGGCAGATGACATGTACAGCTTTATGCTGGGAATCCATGTGTA 300  
880 gaagttagcactgtaagcttctgacacccctccatgaagaagaagaagacactcctt 939  
301 GAGTTATCTCTGCTCACTCATTTTGACACCTCCCTCATTTAGAGAGCAAGCATCTCTT 360  
940 gaggcaaaaacgaagaaccagcaagtcacctataaacttgcataaagataaattt 999  
361 GAGCAAAAACAGCAAGAGACCAGCAAGTCCCTATTAACCTTGTATTAAGTATATTT 420  
1000 gaattatcgctgattcaacatgacttggataagacagcgccttgccttgctgtg 1059  
421 GAATATTCCTGCTGTTTCAACATGTTCTTTGATTAATGATGCGCTTGGCTTGGCTGTG 480  
1060 attacactcttaacatattggaacatgataccttgatataatgacatcattatagg 1119  
481 ATTATCACTCTTACAAATATTTGGAGATGATGCTGATATGATGATCATTTATAGC 540

QY 1120 atgacaaccgaagatcgaatgagatgacgttaccgttcgacgaattagaagaagg 1179  
|||||  
DB 541 ATGACAAACGAGAGATTCAGATGATTCATGATTTACTCTGCGACAAATGAGAAAGGG 600

QY 1180 gt 1181  
11  
DB 601 GT 602

## RESULT 5

AM249865/c 619 bp mRNA EST 07-JAN-2000  
LOCUS  
DEFINITION  
2821628.3prime NIH-MGC.7 Homo sapiens cDNA clone IMAGE:2821628 3',  
mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
AM249865  
AM249865.1 GI:6592858  
EST.  
human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 619)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Other ESTs: 2821628.5prime  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: DCTD/DRP CDNA Library Preparation: Ling  
Hong/Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E.  
Consortium (LLNL) DNA Sequencing by: Berkeley MGC sequencing  
Project Clone distribution: MGC clone distribution information can  
be found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/dbp/image/image.html](http://www.bio.llnl.gov/dbp/image/image.html) Base Calling / Quality  
Scores: PHRED from University of Washington Genome Center. Vector  
Trimming: cross-match from University of Washington Genome Center  
PhRAP suite. Poly-T identification: patchwork.pl from Berkeley  
Drosophila genome project. University of Washington Genome Center:  
<http://www.genome.washington.edu> Polyadenylation: Based upon the  
presence of a XhoI site followed by a run of 14 or more T residues  
at the beginning of the sequence, this cDNA insert was  
polyadenylated.

plate: LHCN7 row: E column: 21  
High quality sequence stop: 504.

## FEATURES

## Source

Location/Qualifiers  
1..619  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2821628"  
/clone\_lib="NIH-MGC.7"  
/tissue\_type="small cell carcinoma"  
/cell\_line="MGC3"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lung; Vector: pOTB7; Site: 1: XhoI, Site 2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCACGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 205 a 111 c 85 g 218 t  
ORIGIN

Query Match 29.1%; Score 591; DB 113; Length 619;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 591; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1442 taacacattagaatagctgctgtatggaagaaagtgacatgaattatgaacaact 1501  
|||||  
DB 597 TAAACATTATGATAGCTGCTGTATGGAAGAAAGTGACATGAATTTATTAAGACAACT 538

OY 1502 taagaagcttaactcttcaacagcatagtggaataatcatatttggtctattgtatcac 1561  
 |||||||  
 DB 537 TACGAATGCTTAACCTCTTTACACAGCATAGTGAAATCATATTTGGCTATTGTATAC 478  
 |||||||  
 OY 1562 tatgaacattcttaattcttaatttgatglaaataactctgaacaagaagaagtt 1621  
 |||||||  
 DB 477 TATGACAAATTTTAAATGCTTATTGATGTAAATTAACCTGAAACAGAGAAAGGT 418  
 |||||||  
 OY 1622 tttaacttagtagagcctaataatagatgtgctataataacogcttagtttggaac 1681  
 |||||||  
 DB 417 TTTTAACTAGAGTACCCCTTAAATATGATGCTTATATATGCTTAGTTTGGAAC 358  
 |||||||  
 OY 1682 tctactagtagaacaagagacagctgtttttaaaccctctctcgaagtttgttgaac 1741  
 |||||||  
 DB 357 TGTATCTGAGTAACAGAGACAGCTGTTTTTAACTCTTCTGCAAGTTTGTGACCTA 298  
 |||||||  
 OY 1742 catgggctaataatgatactaaataactatcatctgaagaagaacagctgtgtg 1801  
 |||||||  
 DB 297 CATGGCTAAATATGATGACTAAATATCTATGATGATGATGATGATGATGATG 238  
 |||||||  
 OY 1802 agtatagatgcttctcatatatacaacaataatccctgaagagacatttgagagatga 1861  
 |||||||  
 DB 237 AGTATATGATGCTTTTCAATATACACAAAATCCCTGAGAGCATTTTGAGGATGA 178  
 |||||||  
 OY 1862 atataaacaattttatctcaagtaacttcccccgtgtgaagtaagtaagttgtgt 1921  
 |||||||  
 DB 177 ATATAAACATTTTATTTTCACTTACCTTTCCCTGCTGATGATGATGATGATG 118  
 |||||||  
 OY 1922 acaattcatctatagaataatgaatgaagtgagtggaatcttcaatttgatggag 1981  
 |||||||  
 DB 117 ACACTTCTATCTATGATGATGATGATGATGATGATGATGATGATGATGATG 58  
 |||||||  
 OY 1982 tggaccaatgtctatacaagatgacaataaagtaagtaagtaagtaagtaagta 2032  
 |||||||  
 DB 57 TCGACCAATGCTATCAAGATGACAAATTAAGTATGATGATGATGATGATG 7  
 |||||||  
 RESULT 6  
 A1091273 586 bp mRNA EST 30-OCT-1998  
 LOCUS 0622B03.X1 Soares.NSF\_F8\_9M\_OT\_PA\_P\_S1 Homo sapiens cDNA clone  
 DEFINITION IMAGE:1651373.3', mRNA sequence.  
 ACCESSION A1091273  
 VERSION A1091273.1 GI:3430332  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 586)  
 NC1-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-remail.nih.gov  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
 Insert length: 600 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 447.  
 Location/Qualifiers  
 1..586  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1651373"  
 /clone\_lib="Soares.NSF\_F8\_9M\_OT\_PA\_P\_S1"  
 /lab\_host="DH10B"  
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
 a modified polylinker; Site: 1: Not 1; Site: 2: Eco RI;  
 Equal amounts of plasmid DNA from five normalized  
 libraries were mixed, and ss circles were made in vitro.  
 Following HAP purification, this DNA was used as tracer in

a subtractive hybridization reaction. The driver was  
 PCR-amplified cDNAs from pools of 5,000 clones made from  
 the same 5 libraries. The pools consisted of the following  
 libraries and clones: Soares NBHP pool 1:  
 309384-310919, 323208-325895 Soares NB2HP pool 1:  
 145032-147335, 147720-148103, 148872-149255, 15002 -  
 150407, 151176-152327 Soares NB2HF-9W pool 1:  
 758280-760583, 772104-774407 Soares NBHPA pool 1:  
 304776-306311, 320136-322823, 326280-326663 Soares NBHP  
 pool 1: 723720-726407, 739080-740999 Subtraction by Bento  
 Soares and M. Patricia Bonaldo.  
 BASE COUNT 199 a 107 c 82 g 198 t  
 ORIGIN  
 Query Match 28.7%; Score 584; DB 16; Length 586;  
 Best local similarity 100.0%; Pred. No. 0;  
 Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1449 ttgaagaagctgctgtatggaanaaagtcagcgaattatlaagacaacttcaagat 1508  
 |||||||  
 DB 586 TTTAGATAGCTGCTGTATGGAATAAGTGCACATGATTTATGACAACTTACGAAT 527  
 |||||||  
 OY 1509 gcttaactcttcaacagcatagtggaataatcatatttgagcattgtatactatgaac 1568  
 |||||||  
 DB 526 GCTTACTCTTTTACACACATGATGATGATGATGATGATGATGATGATGATGATG 467  
 |||||||  
 OY 1569 aattgtaaatgcttaatttgaatgaataaactctgaacaagaagaagaagttttaa 1628  
 |||||||  
 DB 466 AATTTGTAAATGCTTAAATTTGATGATGATGATGATGATGATGATGATGATG 407  
 |||||||  
 OY 1629 ttgaagagcccaataatgaatgagtgctataatcgcctagtttgaagctgatat 1688  
 |||||||  
 DB 406 TTAGAGTACCCCTTAATATGATGATGATGATGATGATGATGATGATGATGATG 347  
 |||||||  
 OY 1689 gagtaacagagagagcgtgtttaaaccctctcgaagttgttgtaacctatagggc 1748  
 |||||||  
 DB 346 GAGTAACAGAGAGAGCTGTTTTTAAACCCCTTCTGCAAGTTTGTGACCTCATGAGGC 287  
 |||||||  
 OY 1749 taatagatatacaataatcatatgataagaaagaactgaacctgtgagatatac 1808  
 |||||||  
 DB 286 TATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 227  
 |||||||  
 OY 1809 agatgcttctcatatatacaacaataatccctgaagagacatttgagagcatgataa 1868  
 |||||||  
 DB 226 AGATGCTTTTCAATATACACAAAATCCCTGAGGACATTTTGAGGACATGATTA 167  
 |||||||  
 OY 1869 acattttatctgaacttcccccgtgtgaagttactatggtttgtgtacaact 1928  
 |||||||  
 DB 166 ACATTTTATTTTCACTTCTTCCCTGCTGATGATGATGATGATGATGATGATG 107  
 |||||||  
 OY 1929 catctatagaataatgaatgagtgagtgaaatcttacttcttctgtgagtgagaca 1988  
 |||||||  
 DB 106 CATTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 47  
 |||||||  
 OY 1989 atgtctatcaagatgacaataaagttatcatgataatccaaa 2032  
 |||||||  
 DB 46 ATCTCTATCAAGAGTGCACAAATTAAGTATGATGATGATGATGATG 3  
 |||||||  
 RESULT 7  
 BF726753 572 bp mRNA EST 05-JAN-2001  
 LOCUS by11d12.Y1 Human Lens cDNA (Un-normalized, unamplified): BY Homo  
 DEFINITION sapiens cDNA clone by11d12 5', mRNA sequence.  
 ACCESSION BF726753  
 VERSION BF726753.1 GI:12042664  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 572)

AUTHORS Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.  
 TITLE NEIBANK: EST analysis and bioinformatics for ocular genomics  
 JOURNAL Invest. Ophthalmol. Vis. Sci. 41 (2000) In press  
 COMMENT Section on Molecular Structure and Function  
 National Eye Institute  
 6/331, NIH, Bethesda, MD 20892-2740, USA  
 Tel: 301 402 3452  
 Fax: 301 496 0078  
 Email: graeme@helix.nih.gov  
 Plate: 11 row: d column: 12

FEATURES  
 source Seq primer: M13RP1 reverse primer (AB1).  
 Location/Qualifiers  
 1..572

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="by11d12"  
 /clone\_lib="Human Lens cDNA (Un-normalized, unamplified):  
 by"  
 /tissue\_type="Lens"  
 /dev\_stage="Adult"  
 /lab\_host="EMDH10B"  
 /note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses  
 from different adults (both approximately 40 years old)  
 together yielded 20ug of total RNA and 150ng mRNA for cDNA  
 library synthesis. A directionally cloned cDNA library in  
 the pCMVSPORT6 vector was constructed at Life Technologies,  
 essentially following the protocols of the Superscript  
 plasmid System full details of which are contained in the  
 manufacturer's instruction manual  
 (http://www.lifetech.com/). First strand synthesis was  
 carried out using a Not I primer-adaptor  
 [5'-PGAGTACTGTCTAGATCGGACGCGCCGCT(15-3') Not I/dlunt  
 end inserts were cloned into the Not I/EcoR V sites in the  
 vector. EST analysis was performed on the unamplified  
 library at the NIH Intramural Sequencing Center (NISC)."  
 BASE COUNT 193 a 78 c 105 g 196 t  
 ORIGIN

Query Match 28.1%; Score 572; DB 168; Length 572;  
 Best Local Similarity 100.0%; Pred. No. 3; Re-293;  
 Matches 572; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1411 atgaattggaatgacgcgctgaagaagaatgtaaacattagaaatgcgtgtatg 1470  
 DB 1 ATGAATTTGGAAATATGACCTGAAGAAATGTAAACATTATGAAATGCTGTTATGG 60  
 QY 1471 aaaaaatgactgaattatagacaacttgcagatgcttaactcttacaacagat 1530  
 DB 61 AAAAAAGTGCATGATTTATTAGCAAACTTACGATGCTTACTTTACACAGCAT 120  
 QY 1531 aggtgaagaatcatattggcgctattgatactgaacaattgtaaatgtcttaattg 1590  
 DB 121 AGGTGAAGAAATCAATTTGGCTATTGTATGACAAATTTGTAATGCTTAAATTTG 180  
 QY 1591 atgtaaatcaacttgaacaagaagaagtttctaactagatgacccataaatatg 1650  
 DB 181 ATGTAAATTAATCTGTAACAGAGAAAAGTTTAACTTAGATGACCTTAATAATATG 240  
 QY 1651 atggtctataatagcgttagtttggaaactgctatcgatgaacagaggaagctgttt 1710  
 DB 241 ATGTCTTAATATATGCTGTGTTGTTGAACTGATCTGATGACAGACAGACGCTTT 300  
 QY 1711 tttaacctcttcgcaagtttgttgcactacatggtgtaataatgataaataact 1770  
 DB 301 TTAAACCTCTCTGCAAGTTTGTGACCTACATGGCTAATATGATCTAATAAATACT 360  
 QY 1771 acattgctcaagaagaagaacagctctgtggagtatataagctgttcttaatacacac 1830  
 DB 361 ACATTGATCTTAAGAAACAACTTACCTTGTGACCTATATAGATGCTTTTCAATTAACAC 420  
 QY 1831 aaaaatccctgagagacatttggagcatatataaacaatttattcagtaacttt 1890

DB 421 AAAAAATCCCTGAGGACATTTTGAGCGCATGATTAACATTTTATTTCACTAACCTTT 480  
 QY 1891 tccccctgtgaagctactatggtttgtgtacaaactcatctatataagaataagtg 1950  
 DB 481 TCCCCCTGTGAAGTACTATGCTGTTGTGGTACAACTTCATTATATGATTAAGTGG 540  
 QY 1951 aagtggtggaattctactcttattatgttgagct 1982  
 DB 541 AAGTGGTGAATTTCTACTTTTATGTTGAGT 572

RESULT 8  
 BF969051  
 LOCUS 954 bp mRNA EST 22-JAN-2001  
 DEFINITION 602269980F1 NIH\_MGC\_84 Homo sapiens cDNA clone IMAGE:4358275 5',  
 mRNA sequence.  
 ACCESSION BF969051 GI:12336266  
 VERSION BF969051.1 GI:12336266  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 954)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 CONTACT: Robert Strausberg, Ph.D.  
 Email: cgabbs@mail.nih.gov  
 Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LMU)  
 Cloning by: Inyele Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LMU at:  
 http://image.llnl.gov  
 Plate: LLM9995 row: P column: 20  
 High quality sequence stop: 652.  
 Location/Qualifiers  
 1..954

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4358275"  
 /clone\_lib="NIH\_MGC\_84"  
 /tissue\_type="adrenal cortex carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site: 1:  
 NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT  
 primed. Average insert size 1.228 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH-MGC Library."  
 BASE COUNT 259 a 156 c 221 g 318 t  
 ORIGIN

Query Match 27.3%; Score 554; DB 172; Length 954;  
 Best Local Similarity 99.8%; Pred. No. 1.5e-283;  
 Matches 604; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 992 ataatttgaattccgtgttttcaacatgtaacttggataatgacgcctggcct 1051  
 DB 1 ATAAATTTGCAATATTCCTGTTTCAACATGATCTTGGATTAATGATGCGCTTGGCCT 60  
 QY 1052 tggcttgatitacccctcaacatatttggaaatgtaactctgataatgatacatca 1111  
 DB 61 TGGCTGTGATTATCCCTCTTACAAATTTTGGAAATGATGATCTGGATATGATCAATCA 120  
 QY 1112 ttataagatgacaacacagaagatcgaatgtaattgaatgtaattgacatgacgaattag 1171  
 DB 121 TTATAGATGACCAACAGAAATTCGAATGATGATGATGATGATGATGATGATGATGAT 180  
 QY 1172 aaaaagggtgtggaatgctgttctgttaataatataatctttagtctgtttaaagta 1231

```

|||||
Db 181 AAAAGGGGTTGGAATGGCTTTGTTAAATATATCTTTAGTGTCTTTAAAGTA 240
Qy 1232 gaagataactatcattatataaaaaaataaattgtcttcttatttctgtgtcc 1291
|||||
Db 241 GATAGTATAGCTTACATTTATATAAAAAAATCAAAATTTGTTCTTATTTTGTGTGTC 300
Qy 1292 tctgtgtcttcttctagatggaattatagatctgaagctgaagccagtggtataatc 1351
|||||
Db 301 TGTGATGTTTTTCTAGAGGAATATATATATATATATATATATATATATATATATAT 360
Qy 1352 caataatgcttgaataatataataataataataataataataataataataataata 1411
|||||
Db 361 CATATATATGCTGTAATATATATATATATATATATATATATATATATATATATATAT 420
Qy 1412 tgaatttggaataatgactggaagaataatgtaaacattgaataatgctgtttatga 1471
|||||
Db 421 TGAATTTTGGAAATATGACCTGGAAGAAATGTAAACATTTAGAAATGCTGCTGTATGA 480
Qy 1472 aaaaagctgactgaattatagacaacttaagaactgactgaactcttcaacaacata 1531
|||||
Db 481 AAAAGTGCATGAAATTTATATGACAACTTAGCAATGCTTAACTTTTATACACAGATA 540
Qy 1532 ggtgaaatcatattggcgtatgtatatacgaacaattgttaagtcttaattga 1591
|||||
Db 541 GGTGAAATCATATATTTGGCTATTTATATATATATATATATATATATATATATATGA 600
Qy 1592 tcttaa 1596
|||||
Db 601 TGTAA 605
|||||

RESULT 9
LOCUS BE894114 564 bp mRNA EST 20-OCT-2000
DEFINITION 601438242F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3923266 5',
mRNA sequence. NIH
ACCESSION BE894114
VERSION BE894114.1 GI:10356157
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 564)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapds-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DMP
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M9759 row: c column: 11
High quality sequence stop: 564.
Location/Qualifiers
1. 564
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: skin; Vector: pCMV-SPORT6; site: 1; NotI;
site 2: SalI; Cloned unidirectionally. primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT 186 a 80 c 105 g 193 t
ORIGIN

```

```

Query Match 27.1%; Score 551; DB 141; Length 564;
Best local similarity 100.0%; Pred. No. 5; Be-282;
Matches 551; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1460 tctgtttagtgaagaagtgacatgaattatattagacaacttgaagtgacttact 1519
Db 1 TGTGTTATGGAAGAAAGTGCACTGAATTTATATGACAACTTACGAATCTTAACTTCT 60
Qy 1520 ttacacagcatagtgtaaaataatatttggcctattgtatctatgaaacaatttgaat 1579
Db 61 TTACACACATAGGTGAAATCATATTTGGCTATTTGTTATATATGAAATTTTGAAT 120
Qy 1580 gctcttaatttgatgaataactcgtgaacagaagaagaaggttttaactgaagtagcc 1639
Db 121 GCTTATATTTATATATATATATATATATATATATATATATATATATATATATATAT 180
Qy 1640 ctaaaatactgagtgcttataataatcgcttagtlltggaaactgtatcagtaacagag 1699
Db 181 CTAAATATATGATGTGCTTATATATATATATATATATATATATATATATATATATAT 240
Qy 1700 gacagctgtttttaaaccctctctgcgaagttgttgacctacatgagcctaataatgata 1759
Db 241 GACAGCTGTTTTTAACTCTTCTTCTGCAAGTTGTTGACCTACATGGCTAATATATGATA 300
Qy 1760 ctaaaatactacatctgacttaagaagaactgacgttggagatataatgacttctc 1819
Db 301 CTAAATATATATATATATATATATATATATATATATATATATATATATATATATATAT 360
Qy 1820 attatagaacaaaatcccttgaaggaacttggagcatgaaataaacaatttatt 1879
Db 361 ATTATATACACAAAATCCCTGAGGACATTTTGAGCATGAAATTAACATTTTATAT 420
Qy 1880 tcaagtaacttccccctgtgtaagtaactatagtttggtaacacttactataga 1939
Db 421 TCAGTAACTTTTCCCTCTGTGAAGTATGATGTTGGTGAACAATTCATTTCTATAGA 480
Qy 1940 atattaaagtgaagtgggaacttacttcttattgttggagtgagcaatgacttcaa 1999
Db 481 ATTTTATAGTGAAGTGGGTGAATTTTACTTTTATGTTGAGTGGACCAATGCTATCA 540
Qy 2000 gaagtgaacaat 2010
Db 541 GAGTGACAAAT 551

RESULT 10
LOCUS AB854119/c 546 bp mRNA EST 31-DEC-1998
DEFINITION aJ71b08.s1 Soares_papathyroid_tumor_NBHPA Homo sapiens cDNA clone
IMAGE:1401879 3', mRNA sequence.
ACCESSION AB854119
VERSION AB854119.1 GI:2941657
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 546)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap/
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapds-remail.nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

```





QY 1742 catyggctaataatgactactaataactacatgactaagaagaactagcctgtgtg 1801  
 |||||  
 Db 298 CATGGGCTAATATGATGACTATAAATCTACATGATCTAAGAAAGAAACAGCGCTTGG 239  
 QY 1802 agtatagatgtcttcttattacacacacaaatccctggagacatttggagcatga 1861  
 |||||  
 Db 238 AGTATATAGATGCTTTTCTATATACACAAAATCCCTGGAGGACATTTTGAAGCATGA 179  
 QY 1862 atataaacaatttatttctcagtaacttcccccgtgtgaagtlactatgtgtgt 1921  
 |||||  
 Db 178 ATATATAAACAATTTTATTTCACTTTCCTCCCTGCTAAGTATGATGATGCTGTGT 119  
 QY 1922 acaactcattcatagaaatlaagtgagagtgagtgagtgagtgagtgagtgag 1981  
 |||||  
 Db 118 ACAACTTATCTTATATGATATTAAGTGAAGGAGGATGATCTACTTTTATGTTGAG 59  
 QY 1982 tggacaatgtctatacaagatgacaataaagtaatgagtgatccaaga 2032  
 |||||  
 Db 58 TGGACCAATGCTATTCACAGTGAACAATAAGTATGATGATGATGATGATGATGATG 8

RESULT 12  
 LOCUS BG403007 996 bp mRNA EST 12-MAR-2001  
 DEFINITION 60241818F1 NIH\_MGC\_93 Homo sapiens cDNA clone IMAGE:4525781 5',  
 mRNA sequence.  
 ACCESSION BG403007  
 VERSION BG403007.1 GI:13296455  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 996)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: gcrabs-remail.nih.gov

Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILML)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/ILML at:  
 http://image.llnl.gov  
 Plate: ILML0432 row: d column: 06  
 High quality sequence stop: 702.

FEATURES  
 Location/Qualifiers  
 1..996

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4525781"  
 /clone\_1ib="NIH\_MGC\_93"  
 /tissue\_type="transitional cell papilloma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: bladder; Vector: pCMV-SPORT6; Site: 1; Ncti;  
 Site: 2; Salt: Cloned unidirectionally; oligo dT primed.  
 Average insert size 1.7 Kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC library."  
 BASE COUNT 279 a 186 c 235 g 296 t  
 ORIGIN

Query Match 26.0%; Score 529; DB 153; Length 996;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-270;  
 Matches 529; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 449 ctaagaacaccctgtctgttctcagtggtcccaatgagaagaagtgatatagttag 508  
 |||||  
 Db 8 CTGAGGAACCTCGTGTCTTTGCAATGGCTCCCACTGAGGAAGAGTATATGTAG 67

QY 509 ggaaggaacatcagtggttgaagaaccttcaagtaaccttgcgcagctcgtlaacg 568  
 |||||  
 Db 68 GGAAGGAACATCTAGTGTGTAAGACCTTTCAGTACCTTGCCACTCCCTTAATGCC 127  
 QY 569 tcttcaagaacactgttctcagtlactacctccctcaaltctctgagtaagaacaatg 628  
 |||||  
 Db 128 TGTTCAGGAACACTGTGTCTGAGTTCACCTCCCTCAATTCCTGAGGAACAATG 187  
 QY 629 aagtgaccgtccttcttcttcaagtgagtgagtgagtgagtgagtgagtgag 688  
 |||||  
 Db 188 AAGTGGACCTGCTCTTCTTCTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 247  
 QY 689 ctcttcaagaacatcagtgagtgagtgagtgagtgagtgagtgagtgagtgag 748  
 |||||  
 Db 248 CTCGTCATAGCATATGAGCAAGGATCATCTCTGATTTATATTCAGTGAAGGAG 307  
 QY 749 gtttgatgaatattggaagtgatgagtgagtgagtgagtgagtgagtgagtg 808  
 |||||  
 Db 308 GTTTGATGAATTTGGGAAGGCTTATGGAAGACTCTGACATTTAGAGATGCTTCT 367  
 QY 809 agatcctgttgcagctctgcacaaagttgcagatgacatgacatgacatgacatg 868  
 |||||  
 Db 368 ACATCTCTGTGACGCTCTGCAAAAGTTGACATGATGATGATGATGATGATGATG 427  
 QY 869 atcagtgatgaagtgatgagtgagtgagtgagtgagtgagtgagtgagtgag 928  
 |||||  
 Db 428 ATGCAAGTGATGAGTGTAGTACGTCAGTCAAGTCAATTTGACCTCTCTTATGAGAGACA 487  
 QY 929 ggaatccttgcagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 977  
 |||||  
 Db 488 GGACTATCTTGAAGCAAAACAGGAAGAAACCAAGCAAGTCTATATA 536

RESULT 13  
 AA427616 521 bp mRNA EST 16-OCT-1997  
 LOCUS AA427616/c  
 DEFINITION zw30c10.s1 Soares ovary tumor NBHOT Homo sapiens cDNA clone  
 IMAGE:770802 3' similar to contains Alu repetitive element; mRNA  
 sequence.  
 ACCESSION AA427616  
 VERSION AA427616.1 GI:2111449  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 521)  
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
 Kucaba, T., Lacy, M., Le, N., Lennon, G., Merriam, J., Martin, J., Moore, B.,  
 Schellenberg, K., Stepien, M., Tan, F., Theisling, B., White, Y., Wylie,  
 T., Waterston, R., and Wilson, R.  
 WashU-Merck EST Project 1997  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu

This clone is available royalty-free through ILML; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -41m13 fwd. ET from Amerisham  
 High quality sequence stop: 409.

FEATURES  
 Location/Qualifiers  
 1..521

/organism="Homo sapiens"  
 /db\_xref="GDB:5980675"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:770802"  
 /clone\_1ib="Soares ovary tumor NBHOT"  
 /sex="Female"  
 /tissue\_type="ovarian tumor"  
 /lab\_host="DH10B (ampicillin resistant)"

/note="Organ: ovary; Vector: p773D (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAGTGGAGCGGCGGCTTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified p773 vector (Pharmacia). Library constructed by Bento Soares and M. Fatima Ronaldo."

BASE COUNT 180 a 96 c 75 g 170 t  
ORIGIN

Query Match 25.6%; Score 521; DB 7; Length 521;  
Best Local Similarity 100.0%; Pred. No. 5,6e-266;  
Matches 521; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1509 gcttaactctttcacagcatagtgaaatactatcttggctatgttactatgaac 1568  
|||||  
Db 521 GCTTAACCTTTTACACAGCATAGTGAATAATCTATTGGCTATTGTATACATGAAC 462  
QY 1569 aattgtaaatgtcttaatttgaatgaatctcgaacaagaagaagttttaa 1628  
|||||  
Db 461 AATTGTAAATGCTTTAAATTTGATGTAATACCTGAAACAAGAAAGTTTAAAC 402  
QY 1629 ttagaatagccctaaataatgatgtgtatataatgccttaagtgtgaactgatat 1688  
|||||  
Db 401 TTAGAGTAGCCCTAAATATATGATGTCTTATATATCCCTTAGTTTGGACCTGATCT 342  
QY 1689 gagtaacagagacagctgtttttaaaccctctctgcaagtttggactacatgagc 1748  
|||||  
Db 341 GAGTACAGAGAGACCTGTTTTTAAACCTTTCGCAAGTTGTGGACCTACATGAGGC 282  
QY 1749 taataagatactaaataactacatgatacagaagaactaactccttggagatata 1808  
|||||  
Db 281 TAAATGTGATCTAAATAATCTACATGATGATTAAGAAAGAACTACCTTGTGAGATAT 222  
QY 1809 agatgctttcatatatacacaataatccctgaaggacatttgaagcatgaataa 1868  
|||||  
Db 221 AGATGCTTTTCAATATATACACAAAAAATCCCTGAGGACATTTTGAGCATGATATAA 162  
QY 1869 aacttttattcagtaactttcccccctgtgaagtctatgattgttggtaacact 1928  
|||||  
Db 161 ACATTTTATTTAGTAATCTTTTCCCTGTGTAATGATATGTTTGGTACACTT 102  
QY 1929 catcttataaataataagtggaagtgtggaattctacttcttattgttggatgacca 1988  
|||||  
Db 101 CATTTATATAGATATATTAAGTGAAGTGGTGAATCTCTATTTTATGTTGAGCTGACCA 42  
QY 1989 atgtctatcaagatgacaaataaagttaattgatttca 2029  
|||||  
Db 41 ATGTCTATCAAGAGTGACAAATAAAGTTAATGATGATGCCA 1

RESULT 14  
AM968992/c 648 bp mRNA EST 01-JUN-2000  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Human; Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
Hedge, P., Qi, R., Abernathy, K., Diarap, S., Gaspar, R., Gay, C., Holt,  
I. E., Saeed, A. I., Sharov, V., Lee, N. H., Yeatman, T. J. and  
Quackenbush, J.

TITLE  
JOURNAL  
COMMENT  
Assessment of gene expression patterns in a model of colon tumor  
metastasis using a 19,200 element cDNA microarray  
Unpublished (2000)  
Contact: John Quackenbush

The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 3528  
Fax: 301 838 0208  
Email: johng@igr.org  
Plate: 259  
Seq primer: Forward

FEATURES  
Source location/Qualifiers  
1..648  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="MAGE resequences, MAGJ"  
/note="Vector: pBluescriptSKm"

BASE COUNT 226 a 124 c 86 g 212 t  
ORIGIN

Query Match 25.5%; Score 518; DB 122; Length 648;  
Best Local Similarity 99.8%; Pred. No. 2.2e-264;  
Matches 568; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1128 ccagaagatlcgaatgattgaattgactctgtccagaattgaaagggttggaaa 1187  
|||||  
Db 603 CCAGAAGATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 544  
QY 1188 ttgactgtttgtttaaataatactttagtgtgtttaaagttagatatacttaca 1247  
|||||  
Db 543 TTGGCTGTTTGTAAATAATATCTTTTACTGTGCTTAAAGTAGATGATGATCTTACA 484  
QY 1248 ttataaaaaaaaataaatttcttcttatttcttatttcttcttcttcttcttcttct 1307  
|||||  
Db 483 TTATATAAAAAATAAATAATTTTGTCTTATTTTGTGCTTAAAGTAGATGATGATCTTACA 424  
QY 1308 agtgaattatagatgtgacgtgaatccactgtgtatagattccataatgttgaat 1367  
|||||  
Db 423 AGTGAAATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 364  
QY 1368 attatgatagcatttaataacatgatttcatcttcttcttcttcttcttcttcttct 1427  
|||||  
Db 363 ATTATGATATAGCATTTATATATATGATGATGATGATGATGATGATGATGATGATGAT 304  
QY 1428 cactgaaagaatgttaaaactttaagaatgctgtgtatgtgaagaagaagtcactgaat 1487  
|||||  
Db 303 CACTGAAAGAATGTAAACATTTAGATATGCTGTGTTATGAAAAAGTGCATGAT 244  
QY 1488 ttattagacaactttagaacttacttcttcttcttcttcttcttcttcttcttcttct 1547  
|||||  
Db 243 TTATTTAGCAAACTTACGATGATGATGATGATGATGATGATGATGATGATGATGAT 184  
QY 1548 gggctatgtatataataataatgttgaatgttcttcttcttcttcttcttcttcttct 1607  
|||||  
Db 183 GGGCTATGTATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 124  
QY 1608 acaagagaagaagttttaaactttagagtagccctaaataalagatgttcttataatag 1667  
|||||  
Db 123 ACAAGAGAAAAGGTTTAACTTAGAGTAGCCCTAAATFATGATGATGATGATGATGAT 64  
QY 1668 cttagtttggacgttctgagtaaca 1696  
|||||  
Db 63 CTGATTTGGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 35

RESULT 15  
BF060652 516 bp mRNA EST 16-OCT-2000  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Human; Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
Hedge, P., Qi, R., Abernathy, K., Diarap, S., Gaspar, R., Gay, C., Holt,  
I. E., Saeed, A. I., Sharov, V., Lee, N. H., Yeatman, T. J. and  
Quackenbush, J.

TITLE  
JOURNAL  
COMMENT  
Assessment of gene expression patterns in a model of colon tumor  
metastasis using a 19,200 element cDNA microarray  
Unpublished (2000)  
Contact: John Quackenbush

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 516)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL  
 Unpublished (1997)  
 CONTACT: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 469.  
 Location/Qualifiers  
 1..516  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="3390702"  
 /clone\_id="Soares\_NSF\_F8\_9M\_OT\_PA\_P\_S1"  
 /lab\_host="DH10B"  
 /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with  
 a modified polylinker; Site 1: Not I; Site 2: Eco RI;  
 Equal amounts of plasmid DNA from five normalized  
 libraries were mixed, and ss circles were made in vitro.  
 Following HAP purification, this DNA was used as tracer in  
 a subtractive hybridization reaction. The driver was  
 PCR-amplified cDNAs from pools of 5,000 clones made from  
 the same 5 libraries. The pools consisted of the following  
 libraries and clones: Soares NBHSF pool 1:  
 309384-310919, 323208-325895 Soares NBHP pool 1:  
 145032-147335, 147720-148103, 148872-149255, 15002 -  
 150407, 151176-152327 Soares NB2HR8-9M pool 1:  
 758280-760583, 772104-774407 Soares NBHPA pool 1:  
 304776-306311, 320136-322823, 326280-326653 Soares NBHOT  
 pool 1: 723720-726407, 739080-740999 Subtraction by Bento  
 Soares and M. Fatima Bonaldo."  
 BASE COUNT 118 a 119 c 125 g 154 t  
 ORIGIN  
 Query Match 25.4%; Score 516; DB 143; Length 516;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-263;  
 Matches 516; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 142 gggaacgaattatattataatcaccagggcgtgttccgaatggaattg 201  
 |||||  
 Db 1 GGGAGCGAGTTAGTATTATAATCACAGGCGTGTGTTCCGAATGGAATG 60  
 Oy 202 cctatacagaagagcgatcccagaagtgtgcatgtccatggtctctctg 261  
 |||||  
 Db 61 CCTATACGAGAGCGGAGATCCAGACGCTGCTCATGTCATGGCTCTCTG 120  
 Oy 262 gaagacctcttgcgaagactcgaatggttaacctttcaccgtccctcga 321  
 |||||  
 Db 121 GAAGACCTTTTGGCGAGACTCGCAGTGGTAACTGTTTCATCGTCCG 180  
 Oy 322 gtcacgtgtatgtgaaggagtgaaacaactgtctctacccccaggaagt 381  
 |||||  
 Db 181 GTCATGTGATGTGAAGGAGTGAACAACTGCTCTACCCCGAGGAGTCA 240  
 Oy 382 taaccttggagaatgagttcccttcttgaactgttgaacattcattcact 441  
 |||||  
 Db 241 TACCTTTGGAGATGAGTCTTTAGTCTTACAGTGTTCGAATTTCACT 300  
 Oy 442 ttttttcttgcgaactccgtgttcttgcagtgtgtcccaaggaggaaga 501  
 |||||  
 Db 301 TTTATTTTCTGAGGAACCTCTGTTGTTTGCATGTGCTCCAGTGAAGA 360  
 Oy 502 atgttaggaaggaactcaatgatttgaagaccttccagtcacttgcgcag 561  
 |||||  
 Db 361 ATGTAGGGAAGCAACTCAGTGTGTAAGACCTTTCAGTACCTTGCACAC 420  
 Oy 562 aatgcctgtttcaagaactcgttctcagttcactccctcaattctcga 621

|||||  
 Db 421 AATGCGCTTTTGAAGAAACCTGCTTCTAGTACACCCCTCAATTTCTGAG 480  
 Oy 622 aacaatgaattacactgtcttcttcttctcgaactg 657  
 |||||  
 Db 481 AACAAATGAATGATGACCTGCTCTTCTTCTTGAACGTG 516  
 RESULT 16  
 A1678997 508 bp mRNA EST 15-DEC-1999  
 LOCUS t60f08.x1 NCI-CGAP\_Gas4 Homo sapiens cDNA clone IMAGE:225463.3,  
 DEFINITION mRNA sequence.  
 ACCESSION A1678997  
 VERSION A1678997.1 GI:4889179  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 508)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL  
 Unpublished (1997)  
 CONTACT: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 www.bio.llnl.gov/bbrp/image/html  
 Insert length: 1155 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 406.  
 Location/Qualifiers  
 1..508  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="225463"  
 /clone\_id="NCI-CGAP\_Gas4"  
 /tissue\_type="poorly differentiated adenocarcinoma with  
 signed ring cell features"  
 /lab\_host="DH10B"  
 /note="Organ: stomach; Vector: pCMV-SPORT6; Site 1: SalI;  
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.69 kb. Life Technologies catalog #: 11549-011"  
 BASE COUNT 173 a 95 c 71 g 169 t  
 ORIGIN  
 Query Match 25.0%; Score 508; DB 23; Length 508;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-259;  
 Matches 508; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1525 cagcatagtgaaatcatatttgggcattgtatcatatgaacaatttgaa 1584  
 |||||  
 Db 508 CAGCATAGTGAATAATCATATTGCGCTATTGTATCTATGAACAAATTTGTAAT 449  
 Oy 1585 aatttgatgaataactctgaacaagaagaagtttctaactagagtagcc 1644  
 |||||  
 Db 448 AATTGATGATTAATACTCGAACAAGAAAAGTTTAACTTAGAGTACCCCT 389  
 Oy 1645 atatgagatgtctatataatcgcttagtatttggaaactatcgtagta 1704  
 |||||  
 Db 388 ATATGAGATGCTCTTATTAATCGCTTAACTGTAATCTAGTACAGAGACG 329  
 Oy 1705 ctgttttcaacctcttcgaagtgtgttgcacacatggcctaataatga 1764  
 |||||

Db 328 CTGTTTTTAAACCCCTCTTCGACAGTTGTTGACCTACATGAGGCTAATATGATCTATAA 269  
 Oy 1765 aatactaatgattgattgataaagaactagccttgtagagatataatgattcattat 1824  
 Db 268 AATACTACATGATTGATTAGAGAAAGAACTAGCCTTGAGATATGATGATGCTTTTCATTTAT 209  
 Oy 1825 acacacaaanaatccctgtaggagacatttgaggcatalaataaacttatttcaat 1884  
 Db 208 ACACACAAANAATCCCTGAGGACATTTTGAGGACATGATATAAATTTTATTTCAGT 149  
 Oy 1885 aactttccccccttgtaagttactatgatttgtagaacttattcattgaataat 1944  
 Db 148 AACTTTTCCCCCTGTGATGATTACTATGTTGTTGGTACACCTTATCTATGAAATATT 89  
 Oy 1945 aagtggaagtgagtgaaatttacttatttgtagtgagtgagacaaatgctataagaagtg 2004  
 Db 88 AAGTGAAGTGGGTGATTTCTACTTTTATGTTGAGTGAGGACCAATGCTATCAAGAGTG 29  
 Oy 2005 acaataaagttaattgattcaccataa 2032  
 Db 28 ACAATAAAGTTAATGATGATTCACAAA 1

RESULT 17  
 A1041810/c 538 bp mRNA EST 30-JUN-1998  
 LOCUS oy34a01.x1 Soares parathyroid tumor\_NbHPA Homo sapiens cDNA clone  
 DEFINITION IMAGE:1667688 3', mRNA sequence.  
 ACCESSION A1041810  
 VERSION A1041810.1 GI:3281004  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 538)  
 NC-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-r@mail.nih.gov  
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo  
 , Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NC-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/BLNC at:  
 www-bio.lhl.gov/bhrp/image/image.html  
 Seq primer: -40m13 fwd. RT from Amersham  
 High quality sequence stop: 447.  
 Location/Qualifiers  
 1. 538  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="1667688"  
 /clone\_id="Soares-parathyroid\_tumor\_NbHPA"  
 /tissue\_type="parathyroid tumor"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: parathyroid gland; Vector: pT73D (Pharmacia  
 RI; 1st strand cDNA was primed with a Not I - Oligo(dT)  
 primer  
 13'-TGTACCAATCTGAAGTGGGAGCGGCGGACCAATTTTTTTTTTTTTTTTTTTTTT  
 TTTT-3'), double-stranded cDNA was size selected, ligated  
 to Eco RI adapters (Pharmacia), sites of a modified pT73  
 cloned into the Not I and Eco RI sites of a modified pT73  
 vector (Pharmacia). Library went through one round of  
 normalization to a Cot = 5. Library constructed by Bento  
 Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid  
 adenomas was kindly provided by Dr. Stephen Marx, National  
 Institute of Diabetes and Digestive and Kidney Diseases,

BASE COUNT 183 a 96 c 79 g 180 t  
 ORIGIN

Query Match 24.9%; Score 507; DB 15; Length 538;  
 Best Local Similarity 100.0%; Pred. No. 1,6e-258;  
 Matches 507; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1497 aaacttaagaatgcttaactccttcttaacagcagtagtgaaatcatattggcgtatg 1556  
 Db 538 AACTTTAGAAATGCTTAACCTTTTACACAGCATGAGTGAAATCATATTGGGCTATTG 479  
 Oy 1557 tatactatgaacaattgtaaatgcttaattgtagtaataactctgaacaagaagaa 1616  
 Db 478 TATACTATGAAACATTTGTAATGCTTATTGATGATTAATACCTGAAACAGAGAA 419  
 Oy 1617 aaggttttacttaagtagagccttaataataatgtagtgctatataatcgttagtt 1676  
 Db 418 AAGTTTTTAACTTAGAGTAGCCCTAAATATGATGCTTATATATGCTTAGTTTT 359  
 Oy 1677 ggaactgatactgagtaacaagagagagcgttttttaaccctctcgaagttgctg 1736  
 Db 358 GGAAGTATGATCTGATTAACAGAGAGAGAGCTGTTTTTAACCTCTCTCGCAAGTTTGTG 299  
 Oy 1737 actacatgaggtcctaataatgatactaaataactacatgactaagaagaactagcct 1796  
 Db 298 ACTTACATGAGGCTTAATATGATTAATAAATACTATGATCTTAAGAAACATAGCCT 239  
 Oy 1797 tctgagatataatgagcttctcaltatcacacaanaatccctgagagagactttgag 1856  
 Db 238 TCTGAGATATATGATGATGCTTTTCAATATATACACAAATAATCCGAGAGACATTTGAG 179  
 Oy 1857 catgaataaataacatttttattcaglaactttcccccgtgtagtactatgctt 1916  
 Db 178 CATGAATATTAACATTTTATTTCAGTAACCTTTCCCTGTTAGTACTATGCTTT 119  
 Oy 1917 gtgtgaacttattcattatagataatgaatgaatgagtgagtgagtaacttattgct 1976  
 Db 118 GTGTGAACACTTCATCTATGAAATATTAAGTGAAGGAGGATGATCTTATTATGT 59  
 Oy 1977 tggagtgagcaatgctatacaagct 2003  
 Db 58 TGGAGTGACCAATGCTATCAAGAT 32

RESULT 18  
 BE079600/c 579 bp mRNA EST 12-JUN-2000  
 LOCUS R05-BT0624-180400-014-D03 BT0624 Homo sapiens cDNA, mRNA sequence.  
 DEFINITION BE079600  
 ACCESSION BE079600.1 GI:8469881  
 VERSION  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 579)  
 Dias Neto E., Garcia Correa R., Verjowski-Almeida S., Briones M.R.,  
 Nagai M.A., da Silva M. Jr., Zago M.A., Bordin S., Costa F.F.,  
 Goldman G.H., Carvalho A.F., Matsukuma A., Bala G.S., Simpson D.H.,  
 Brunstein A., de Oliveira P.S., Bucher P., Jongeneel C.V., O'Hare  
 M.J., Soares F., Brentani R.R., Reis L.F., de Souza S.J. and  
 Simpson A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

TITLE  
 JOURNAL  
 MEDLINE  
 COMMENT  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil













www.bio.lnl.gov/bbrp/image/image.html  
Seq primer: -40UP from Gibco  
High quality sequence stop: 422.

## FEATURES

## source

Location/Qualifiers

1. .478  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2621008"  
/tissue\_type="medulloblastoma"  
/lab\_host="DH10B (phage resistant)"  
/note="Organ: brain; Vector: p773D-pac (Pharmacia) with a modified polylinker. 1st strand cDNA was prepared from medulloblastoma tumor tissue, and was then primed with a Not I - Oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. This library is normalized. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 109 a 111 c 117 g 141 t  
ORIGIN

Query Match 23.5%; Score 478; DB 112; Length 478;  
Best Local Similarity 100.0%; Pred. No. 4.5e-243;

Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 gggagcaggtttagtataataaccacaggtctgtgttttccgaaatggaatg 201  
DB 1 GGGAGCGAGTTTACTATATTAATAATCACAGAGGCTGTTTTCGAAATGGAATG 60  
QY 202 cctataccagagagcgagatccagacgctgcatgtgcatatgcatggtctctgtga 261  
DB 61 CCTATACCGAGAGAGGAGGATCCACAGCTGGCTGCATGTCCATGGCTTCTGTGAAA 120  
QY 262 gaaagccttctggccagagactgcagtggaacttcttcctcctcgtggctacc 321  
DB 121 GAAGACCTTTTGGCCAGAGACTGCCAGTGGGTAACCTTTTTCATCTGCTCGGAGTACC 180  
QY 322 gtcacgtgtagtgaagagagtgacaacacgtgcttaacccacagcagtgcatitcg 381  
DB 181 GTCATGCTGATGCTGAAGGAGTGAACAACTGGCTACCCAGCAGCATTTGCG 240  
QY 382 tacccttggagagatcagatcctttagcttgaagagtgtagcaatctcatcacc 441  
DB 241 TACCTTTGGAGATGCACTTCCTTTAGCTTGAAGAGTGGCAATTCATTCACCTCC 300  
QY 442 ttaatttcgagaaactcctggtgttgcagttggtctccagtgagaaagatgta 501  
DB 301 TTATTTCTGAGAACTCTGTTGTTTGCAGTTGGCTCCAGTGAAGAAAGATGTA 360  
QY 502 atgttaggagagcaactcagttgtgaagaccttcagtcaccttgccagctcgt 561  
DB 361 ATGGTAGGAGCAAACTCAGTGTGTTGAAGACCTTTCAGTACCTTGGCCAGCTCGT 420  
QY 562 aatgcctgttttaagaaaactcgttctcagttcaatccccccaattctcgtgta 619  
DB 421 AATGCCCTGTTTCAAGAAACTCTGTTCTCAGTCACTCCCTCAATTCCTCGATA 478

RESULT 27  
AM513611 473 bp mRNA EST 03-MAR-2000  
LOCUS AM513611  
DEFINITION x047c11.x1 NCI\_CGAP-Ut1 Homo sapiens cDNA clone IMAGE:2707124 3',  
mRNA sequence.  
ACCESSION AM513611  
VERSION AM513611.1 GI:7151689  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 473)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-r@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[image.llnl.gov/image/html/iresources.shtml](http://image.llnl.gov/image/html/iresources.shtml)

Possible reversed clone: polyT not found  
Seq primer: -40UP from Gibco  
High quality sequence stop: 410.

## FEATURES

## source

Location/Qualifiers

1. .473  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2707124"  
/tissue\_type="well-differentiated endometrial adenocarcinoma, 7 pooled tumors"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Life Technologies catalog #: 11538-014"

BASE COUNT 159 a 86 c 66 g 162 t  
ORIGIN

Query Match 22.9%; Score 465; DB 117; Length 473;  
Best Local Similarity 100.0%; Pred. No. 3.8e-236;

Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1568 caatttgtaattccttaatttgaatgaataacctgaacagagaaagttttta 1627  
DB 473 CAATTGTAAATGCTGTAATTTGATGTAATACCTGTAACAGAGAAAGTTTAA 414  
QY 1628 cttagagtaacctaaataatgagtgcttataataatcgtttagtttgaaactgtatc 1687  
DB 413 CTTAGAGTACCCCTAAATATGAGATGCTTATATATCGCTTAGTGGAACTGTATC 354  
QY 1688 tgaatlaacagagacagctgttllttaaccctctcgaagttgttgaactacatgg 1747  
DB 353 TGAGTAAACAGAGACGCTGTTTAAACCTCTTCTGCAAGTGTGTGACCTACATGG 294  
QY 1748 ctaatatgagtaactaaataactcatatgaataagaagaactagccttgtagatata 1807  
DB 293 CTAATATGATGATCTAAATATCTACATGATCTAAGAGAAACTAGCTTGGGAGATA 234  
QY 1808 tagatgcttcaatlaacacacaaaatccctggagagacatttgaagcatgaatata 1867  
DB 233 TAGATGCTTTCAATATACACAAAAATCCCTGAGGACATTTTGAGCATGAAATTA 174  
QY 1868 aacatttatttcagtaacttccctcctgtgtaagttaactatggttggtaaac 1927  
DB 173 AACATTTTATTTAGTAACTTTTCCCTGTGTAAGTATGTTGTTGTGTAACAAT 114  
QY 1928 tcatctclataagataatgaatgagagtggaatcttaacttcttgaatgagagacc 1987  
DB 113 TCATTTATAGAAATATTAAGTGAAGTGGTGAATTTACTTTTATGTTGAGTGAAC 54  
QY 1988 aatgtctatcaagagtgcaaatgaatgaatgatgatccaaa 2032  
DB 53 AATGCTATCAAGAGTGAACAATTAAGTTAATGATGATTCACAAA 9

RESULT 28



Db 284 AATATGATCTACTAAATCTACTATTGATTAAGAGAACTAGCTTGAGAGTATATA 225  
 Oy 1810 gatgctttctatatacacaataatccctggaggaactttggagcagaatataaa 1869  
 Db 224 GATGCTTTCTTATACACAAAATCCCTGAGGAGCACTTTGGCAGATATATAA 165  
 Oy 1870 cattiattcagtaactttcccttgtaagtaactaactgattggtgtaacttc 1929  
 Db 164 CATTATTATTCAGTAACCTTTCCCTGCTGTAAGTACTATGATGCTGTAACACTTC 105  
 Oy 1930 attcagaatatttaagtggaagtgggaatcacttttaattggtgaagtgccaa 1989  
 Db 104 ATTCTATGATATTTAAGTGAAGAGGCTGATTTCTTTTATGTTGAGTGAACAA 45  
 Oy 1990 tgcatacaagatgcaataaagtaataatgaltcccaaa 2032  
 Db 44 TGTCTATCAGAGTGAACAAATTAAGTAAATGATGATTCACAAA 2

RESULT 30  
 AA17296/c 466 bp mRNA EST 09-NOV-1997  
 LOCUS zui8c04.s1 Soares\_NhMpu\_S1 Homo sapiens cDNA clone IMAGE:738342  
 DEFINITION 3', mRNA sequence.  
 ACCESSION AA17296  
 VERSION AA17296.1 GI:2077547  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 466)  
 AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
 Kitzman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin  
 White,Y., Wyllie,T., Waterston,R. and Wilson,R.  
 TITLE WashU-NCI human EST Project  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Wilson RK  
 Washington University School of Medicine  
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: estewatson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (InfoImage.lnl.gov) for further information.  
 Insert Length: 847 Std Error: 0.00  
 Seq primer: -41m3 fwd. ET from Amersham  
 High quality sequence stop: 375.  
 Location/Qualifiers  
 1..466  
 /organism="Homo sapiens"  
 /db\_xref="GDB:5945950"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:738342"  
 /clone\_id="Soares\_NhMpu\_S1"  
 /tissue\_type="Pooled human melanocyte, fetal heart, and  
 pregnant uterus"  
 /note="Organ: mixed (see below); Vector: pUT73P-Pac  
 (Pharmacia) with a modified polylinker; Site\_1: Not 1;  
 Site\_2: Eco RI; Equal amounts of plasmid DNA from three  
 normalized libraries (melanocyte 2NDM, pregnant uterus  
 NDHpu, and fetal heart NDH19M) were mixed, and ss circles  
 were made in vitro. Following HAP purification, this DNA  
 was used as tracer in a subtractive hybridization  
 reaction. The driver was PCR-amplified cDNAs from pools of  
 5,000 clones made from the same 3 libraries. The pools  
 consisted of I.M.A.G.E. clones 260232-265223,  
 340488-345479, and 484488-489479."

BASE COUNT  
 ORIGIN

150 a 96 c 73 g 147 t

Query Match 22.8%, Score 463; DB 7; Length 466;  
 Best Local Similarity 100.0%; Pred. No. 4,4e-235;  
 Matches 463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 816 tctgacgctctgcaaaagttgacagatgacatgacagctcttattggtggaatgagt 875  
 Db 466 tcttgacgctctgcaaaagttgacagatgacatgacagctcttattggtggaatgagt 407  
 Oy 876 gttgagtttgatcgtgcaagttgacatgacatgacatgacatgacatgacatgacat 935  
 Db 406 GGTAGATTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 347  
 Oy 936 ccttgagcaaaaacagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaag 995  
 Db 346 CCTTGAGCAAAAACGAG 287  
 Oy 996 ttgtgaatattcgtggtgttccaacatgtaactttgataatgatactgctgctgctgctgc 1055  
 Db 286 TTTTGAAATATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 227  
 Oy 1056 tctgataatcactcttcaaatatttgacaatgatactcctgataatgataatcattta 1115  
 Db 226 TGTGATTATACCTCTTACATATTTTGAAACATGATCTCTGATATGATATGATATTTTA 167  
 Oy 1116 tagaagcaaacagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 1175  
 Db 166 TAGGATGACAAACAGAGATTCGAAATGGAATGTTACCTGTCGAGAAATTTGAGAAA 107  
 Oy 1176 gagggttggaatgctgctgttggtaaatatcctttagtgcttgaagtagata 1235  
 Db 106 GGGGGTTGGAAATGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 47  
 Oy 1236 gtaacttcaattataaaaaaatacaattgtctctta 1278  
 Db 46 GTATCTTTACATTTATATAAAAAATCAATTTTGTCTTTTA 4

RESULT 31  
 AM964454 546 bp mRNA EST 01-JUN-2000  
 LOCUS EST376527 MAGE resequences, MAGH Homo sapiens cDNA, mRNA sequence.  
 ACCESSION AM964454  
 VERSION AM964454.1 GI:8154290  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 546)  
 AUTHORS Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspari,R., Gay,C., Holt  
 J.E., Saeed,A.T., Sharov,V., Lee,N.H., Yeatman,T.J. and  
 Quackenbush,J.  
 TITLE Assessment of gene expression patterns in a model of colon tumor  
 metastasis using a 19,200 element cDNA microarray  
 JOURNAL Unpublished (2000)  
 COMMENT The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 3528  
 Fax: 301 838 0208  
 Email: johnd@tigr.org  
 Plate: 201  
 Seq primer: Reverse.  
 Location/Qualifiers  
 1..546  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_id="MAGE resequences, MAGH"  
 /note="Vector: pBluescriptSKm"

BASE COUNT 192 a 62 c 99 g 193 t  
 ORIGIN





Db 361 ATGTGATGGAGGAGCAACTCAGTGTGTTGAAGACCTTTGACCTTGGCCAGCTCCGT 420  
 Oy 562 aatgcctgtttcaagaactctgttccag 593  
 ||||||||||||||||||||||||||||  
 Db 421 AATGCGCTGTTCAAGAAACTGCTTCTCAG 452

RESULT 35  
 AMO24459 501 bp mRNA EST 09-MAR-2000  
 LOCUS wu76f06.x1 NCI\_CGAP\_kid3 Homo sapiens cDNA clone IMAGE:2525987 3',  
 DEFINITION mRNA sequence.  
 ACCESSION AMO24459  
 VERSION AMO24459.1 GI:5877989  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 501)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-remail.nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bdrrp/image/image.html  
 Insert Length: 644 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 425.  
 Location/Qualifiers

FEATURES  
 source 1..501

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2525987"  
 /clone\_id="NCI\_CGAP\_kid3"  
 /lab\_host="DH10B"

/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with  
 a modified polylinker. Site.1: Not I; Site.2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer,  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not  
 I and Eco RI sites of the modified pT7T3 vector. mRNA  
 source: 2 pooled kidneys. Library went through one round  
 of normalization. Library constructed by Bento Soares and  
 M. Fatima Bonaldo."

BASE COUNT 115 a 116 c 123 g 147 t

ORIGIN

Query Match 22.1%; Score 450; DB 110; Length 501;  
 Best Local Similarity 99.8%; Pred. No. 3.7e-228;  
 Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 142 gggagcaggttctgtatataaacacagggtctgtgtttccgaatgaattg 201  
 ||||||||||||||||||||||||||||  
 Db 1 GGGAGCAGGTTTGTATATTTAAATCACAGGCTCTGTTTCCGAATGGAATTGG 60  
 ||||||||||||||||||||||||||||  
 Oy 202 cctatacagagagcggatccacagcgtgtgcatgtctcatgaggtctcttga 261  
 ||||||||||||||||||||||||||||  
 Db 61 CCTATACAGAGAGCGGATCCACAGCTGCTGCAATGTCATGCGCTTCTCTGAAA 120  
 ||||||||||||||||||||||||||||  
 Oy 262 gaagaccttttggcaggaactcagatggttaacctgttcatcgtctccgaggtacc 321  
 ||||||||||||||||||||||||||||  
 Db 121 GAAAGACCTTTTGGCGCAGACTCCAGATGGGTACCTTTTCATCGTCTCGGGGTACC 180

Oy 322 gtaatgagatagtggaaggagtgaaacaactggtctacccccagcagtgatcttg 381  
 ||||||||||||||||||||||||||||  
 Db 181 GTCATGTGATGTTGAAGGAGTGAACAACTGGCTCTACCCAGCAGTGTATTTCG 240  
 ||||||||||||||||||||||||||||  
 Oy 382 taaccttggaagaatgcagttcttcttgacttgacagtgltgcaatcattcactcc 441  
 ||||||||||||||||||||||||||||  
 Db 241 TACCTTTTGGAGATTCGAGTTCCTTTTACTCTGTGACAGTGTGCAATTCATCTC 300  
 ||||||||||||||||||||||||||||  
 Oy 442 ttattcttgaggaactcctgtgtgttgcagttggtctccacgtgaggaagtgat 501  
 ||||||||||||||||||||||||||||  
 Db 301 TTATATTTTCGAGAAACCTCTGTTGTCAGTTGCTCCAGTAGGAAAGAGTAT 360  
 ||||||||||||||||||||||||||||  
 Oy 502 atgtgaggaagcaaacactcagtggttgaagaccttcaagtcacacttgccagctccgt 561  
 ||||||||||||||||||||||||||||  
 Db 361 ATGTGAGGAGGAGCAAACTCAGTGTGTTGAAGACCTTTGACGCTTGGCCAGCTCCGT 420  
 ||||||||||||||||||||||||||||  
 Oy 562 aatgcctgtttcaagaactctgttccagatctacccctcaattcttgagg 621  
 ||||||||||||||||||||||||||||  
 Db 421 AATGCGCTGTTCAAGATMACCTGTTCTCAGTTCCATCCCTCAATTCTTGAGTAGG 480  
 ||||||||||||||||||||||||||||  
 Oy 622 aacaatgaagttgacctgtc 642  
 ||||||||||||||||||||||||  
 Db 481 AACATGAGTGTGACCTGCTC 501

RESULT 36  
 AMO03884 448 bp mRNA EST 08-SEP-1999  
 LOCUS w61f04.x1 NCI\_CGAP\_Brn23 Homo sapiens cDNA clone IMAGE:2501695 3',  
 DEFINITION mRNA sequence.  
 ACCESSION AMO03884  
 VERSION AMO03884.1 GI:5850800  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 448)

NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute / National Institute of Neurological  
 Disorders and Stroke, Brain Tumor Genome Anatomy Project

JOURNAL (CGAP/BRGAP), Tumor Gene Index

COMMENT Unpublished (1998)

Contact: Robert Strausberg, Ph.D.

Email: cgaps-remail.nih.gov

Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,  
 Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaldo, Ph.D.

DNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bdrrp/image/image.html  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 445.  
 Location/Qualifiers

FEATURES  
 source 1..448

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2501695"  
 /clone\_id="NCI\_CGAP\_Brn23"  
 /tissue\_type="glioblastoma (pooled)"  
 /lab\_host="DH10B"

/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a  
 modified polylinker. Site.1: Not I; Site.2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5'  
 TGTTCACATCTGATGAGGAGCGACGCGCAATCTTTTCTTTTCTTTTCTTTT  
 T 3']]; double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified pT7T3 vector.  
 Library is normalized, and was constructed by Bento  
 Soares and M. Fatima Bonaldo."





Email: c9apbs-remail.nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: MCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: www-bio.lnl.gov/bhrp/image/image.html  
 Insert Length: 551 Std Error: 0.00  
 Seq primer: -400P from Glibco  
 High quality sequence stop: 444.  
 Location/Qualifiers

## FEATURES

1..451  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1909035"  
 /clone\_lib="MCI-CGAP\_L45"  
 /tissue\_type="carcinoid"  
 /lab\_host="DH10B"  
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker. 1st strand cDNA was prepared from neuroendocrine lung carcinoid, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldi."

BASE COUNT 153 a 84 c 65 g 149 t  
 ORIGIN

Query Match 22.0%; Score 447; DB 18; Length 451;  
 Best local similarity 100.0%; Pred. No. 1.5e-226;  
 Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1585 aattgtagtaataaactctggaagaagaagaagtttlaacttagtagccctaa 1644  
 |||||||  
 Db 447 AATTGATGATTAATTAATCTGGAACAGAGAAAGTTTAACTTAGAGAGCCCTAA 388  
 |||||||  
 Oy 1645 atagtagtgcataataatcgcttagtttggaactatcgtagtaacagaagacag 1704  
 |||||||  
 Db 387 AATAGGATGCTTATATATGCTTACTTTGGAACTGATCTGAGTAACAGAGACAG 328  
 |||||||  
 Oy 1705 cgttttttaacctctgcaagttgttgtagcttaacatgagcaatagatactaa 1764  
 |||||||  
 Db 327 CTGTTTAACTCTCTCGCAAGTTTGTGACCTAATGAGGCTAATATGATACATA 268  
 |||||||  
 Oy 1765 aatactacattgataagaagaactagcctgtgtagatataatagatgcttcat 1824  
 |||||||  
 Db 267 AATACTACATTGATCTAAGAGAACTAGCCCTTGAGATATATGATGCTTTTCATAT 208  
 |||||||  
 Oy 1825 acacacaaatccctgaaggaacatttgaagcagatataaactatllatcaat 1884  
 |||||||  
 Db 207 AACACAAATAATCCCTGAGGACATTTTGAAGCATGAAATATAAATTTTATTTCAG 148  
 |||||||  
 Oy 1885 aacttccccctgtgaagtaactatggttggtaacaaactatcctatagaat 1944  
 |||||||  
 Db 147 AACTTTTCCCTGTGTAAGTACTATGATGTTGGTACAACTTATCTATTAATATT 88  
 |||||||  
 Oy 1945 aatgaagtagtgaattcctactatcttattgttgaatgagcaatgctatcaagag 2004  
 |||||||  
 Db 87 AATGGAAGTGGGATGATCTACTTTTATGTGTGATGAGCAATGCTATCAAGAGTG 28  
 |||||||  
 Oy 2005 acaataaagttaagtatgattccaa 2031  
 |||||||  
 Db 27 ACAATAAAGTTAATGATGATTCCAAA 1

IMAGE:2482454 5', mRNA sequence.  
 AM020149  
 VERSION AM020149.1 GI:5873679  
 EST.  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 human.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS  
 1 (bases 1 to 453)  
 Robertson N.G., Khetarpal U., Gutierrez-Espeleta G.A., Bieber F.R. and Morton C.C.  
 Isolation of novel and known genes from a human fetal cochlear cDNA library using subtractive hybridization and differential screening  
 Genomics 23, 42-50 (1994)  
 95130111  
 JOURNAL  
 MEDLINE  
 COMMENT  
 Contact: Morton, C. C.  
 Departments of Pathology and Obstetrics, Gynecology and Reproductive Biology  
 Brigham and Women's Hospital  
 75 Francis Street, Harvard Medical School, Boston, MA 02115, USA  
 Tel: 617 732 7980  
 Fax: 617 738 6996  
 Email: comoronedics.bwh.harvard.edu  
 DNA sequencing and analyses were performed by National Institutes of Health Intramural Sequencing Center (NISC; see http://www.nisc.nih.gov).  
 This clone is available royalty-free through LINL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.  
 Plate: LLAM6177 row: A column: 15  
 Seq primer: M13RP1 reverse primer (ABI).

## FEATURES

1..453  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2482454"  
 /clone\_lib="Morton Fetal Cochlea"  
 /tissue\_type="cochlea"  
 /lab\_host="16-22 week fetus"  
 /dev\_stage="16-22 week fetus"  
 /note="Organ: ear; Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Reference: Genomics 23, 42-50 (1994) cloned unidirectionally. Primer: Oligo dt. Fetal cochlea, normal. 37% of inserts <0.5 kb, 56% 0.5-1.0 kb, 7% >1 kb. Uni-ZAP XR Vector. Library constructed by N. Robertson, C. Morton. -5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."

BASE COUNT 159 a 61 c 80 g 153 t  
 ORIGIN

Query Match 21.9%; Score 446; DB 110; Length 453;  
 Best local similarity 100.0%; Pred. No. 5e-226;  
 Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1382 atttaataacttgattactatccttgaatgaattggaataatgacagaaagaatg 1441  
 |||||||  
 Db 8 ATTATATACATTTATTTCTTCTGTTAATGAATTTGGAAATATGCACTGAAGAAATG 67  
 |||||||  
 Oy 1442 taaacattggaatagcgtgctgtaagaaagaatgcaactgaattatagacaaact 1501  
 |||||||  
 Db 68 TAAACATTTTGAATACCTCTGTTATGAAAAAAGTCAGCAATTTTATTAGCAAACT 127  
 |||||||  
 Oy 1502 taagaatgctaactctcttaccacagcagatagtaaaatcatatctggagctatgatac 1561  
 |||||||  
 Db 128 TAGAGATGCTTACTTTCTTACACAGCATAGCTGAATAATCATTTTGGCATATGTATAC 187  
 |||||||  
 Oy 1562 tatgaacaattgttaaatgtcttaattgtagtaataactctgaaacaagaagaagt 1621  
 |||||||  
 Db 188 TATGAACAATTTGTAAATGCTTAATTTGATGTAAATCTCTGAACAAAGAAAGGT 247  
 |||||||  
 Oy 1622 tttaacttaagtagccctaaataatagatgcttataataatgcttagttggaac 1681  
 |||||||  
 Db 248 TTTTAACTTAGAGTAGCCCTAAATATATGATGATGCTTATATATGCTTATTGGAAC 307

QY 1682 tgcatacgaagaaagagacgcgtcttttaaccctcttcgaagttgttgaccta 1741  
 |||||  
 Db 308 tggatctgagtaaacagagacgctgttttttaaccctcttcgaagttgttgaccta 367  
 |||||  
 QY 1742 cctggagcaataatgatacctaaataactaactatgatacgaagaaacagcctgtgtg 1801  
 |||||  
 Db 368 cagggcgttaataatgatacctaaataactaactatgatacgaagaaacagcctgtgtg 427  
 |||||  
 QY 1802 agtatatagatgctcttcattatata 1827  
 |||||  
 Db 428 agtatatagatgctcttcattatata 453

## RESULT 40

AA843850 495 bp mRNA EST 04-MAR-1998  
 LOCUS a118906.s1 Soares-parathyroid\_tumor\_MbHPA Homo sapiens cDNA clone  
 DEFINITION IMAGE:1390714 3', mRNA sequence.

ACCESSION AA843850  
 VERSION AA843850.1 GI:2930301  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCICGAP http://www.ncbi.nlm.nih.gov/ncicgap.

REFERENCE 1 (bases 1 to 495)  
 AUTHORS NCI-CGAP  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapdb-r@mail.nih.gov  
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo  
 , Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 www-bio.llnl.gov/bbtp/image/image.html  
 Seq primer: 40m13 fwd. RT from Amersham  
 High quality sequence stop: 467.

## FEATURES

source  
 1..495  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1390714"  
 /clone\_lib="Soares-parathyroid\_tumor\_MbHPA"  
 /tissue\_type="parathyroid tumor"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: parathyroid gland; Vector: pTZ19 (Pharmacia)  
 with a modified polylinker; Site 1: Not I; Site 2: Eco  
 RI; 1st strand cDNA was primed with a Not I - oligo(dT)  
 primer  
 [5'-TGTCACATCTGAGTGGAGCGCGCCACCAATTTTGTGTGTGTGT  
 TTTT-3'], double-stranded cDNA was size selected, ligated  
 to Eco RI adapters (Pharmacia), digested with Not I and  
 cloned into the Not I and Eco RI sites of a modified pTZ19  
 vector (Pharmacia). Library went through one round of  
 normalization to a Cot = 5. Library constructed by Bento  
 Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid  
 adenomas was kindly provided by Dr. Stephen Marx, National  
 Institute of Diabetes and Digestive and Kidney Diseases,  
 NIH."  
 BASE COUNT 166 a 66 c 90 g 173 t  
 ORIGIN

Query Match 21.8%; Score 444; DB 12; Length 495;  
 Best Local Similarity 99.8%; Pred. No. 5-ge-225;  
 Matches 494; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1285 gtgtcctgtgattgttttctagagatattatagattgacgtgaaatccactgtgta 1344  
 |||||  
 Db 1 gtgtcctgtgattgttttctagagatattatagattgacgtgaaatccactgtgta 60  
 |||||  
 QY 1345 tgaattcacaataatgcttgaatattatagatagcattatgaattgattc 1404  
 |||||  
 Db 61 tgaattcacaataatgcttgaatattatagatagcattatgaattgattc 120  
 |||||  
 QY 1405 tgttaataatgattggaataatgacatgacatggaagaatgtaaaacttgaatgctgtg 1464  
 |||||  
 Db 121 tgttaataatgattggaataatgacatgacatggaagaatgtaaaacttgaatgctgtg 180  
 |||||  
 QY 1465 ttatggaanaagtcacatgaaattatagacaacttgaatgcttgaatc 1524  
 |||||  
 Db 181 ttatggaanaagtcacatgaaattatagacaacttgaatgcttgaatc 240  
 |||||  
 QY 1525 cagcaatggaatcacaatcatttggtgctatgatacgaatgtaaaacttgaatgctgtg 1584  
 |||||  
 Db 241 cagcaatggaatcacaatcatttggtgctatgatacgaatgtaaaacttgaatgctgtg 300  
 |||||  
 QY 1585 aatttgatgtaataatcctgaaacagagaagaagtttgaacttagagtgccctaa 1644  
 |||||  
 Db 301 aatttgatgtaataatcctgaaacagagaagaagtttgaacttagagtgccctaa 360  
 |||||  
 QY 1645 atattgattgctatataatcgttgaattggaactgatacgaatgtaaaacttgaatgctgtg 1704  
 |||||  
 Db 361 atattgattgctatataatcgttgaattggaactgatacgaatgtaaaacttgaatgctgtg 420  
 |||||  
 QY 1705 ctgtttttaaaccctcttcgaagttgttgacatacgaatgtaaaacttgaatgctgtg 1764  
 |||||  
 Db 421 ctgtttttaaaccctcttcgaagttgttgacatacgaatgtaaaacttgaatgctgtg 480  
 |||||  
 QY 1765 aatctacattgac 1779  
 |||||  
 Db 481 aatctacattgac 495

## RESULT 41

W58242/c 460 bp mRNA EST 15-OCT-1996  
 LOCUS zdl8c12.s1 Soares\_fetal\_heart\_NbH19M Homo sapiens cDNA clone  
 DEFINITION IMAGE:341014 3', mRNA sequence.

ACCESSION W58242  
 VERSION W58242.1 GI:1364974  
 KEYWORDS EST.

## SOURCE

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

## REFERENCE

AUTHORS 1 (bases 1 to 460)  
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman  
 M., Holtzman, M., Kucaba, T., Le, M., Lennon, G., Marx, S., Parsons, J.,  
 Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston  
 R., Williamson, A., Wohlmann, P. and Wilson, R.  
 The WashU-Merck EST Project  
 Unpublished (1995)  
 CONTACT: Wilson RK  
 Washington University School of Medicine  
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 1256 Std Error: 0.00  
 Seq primer: mob.REGA+RT  
 High quality sequence stop: 386.

## FEATURES

source  
 1..460  
 /organism="Homo sapiens"  
 /db\_xref="GDB:1265389"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:341014"

/clone\_lib="Soares\_fetal\_heart\_NBH19W"  
 /sex="unknown"  
 /dev\_stage="19 weeks"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: heart; Vector: pT73D (Pharmacia) with a modified polylinker; Site: 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I oligo(dT) primer [5' TTTTCCATCTGAGTGGAGCGCCGACATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed from the same focus as the fetal lung library, Soares fetal lung NBH19W."

BASE COUNT 155 a 86 c 65 g 154 t  
 ORIGIN

Query Match 21.8%; Score 443; DB 190; Length 460;  
 Best Local Similarity 100.0%; Pred. No. 2e-224;

Matches 443; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1590 gatgtaataactctgaacagaagaagttttaaacttagagtagccctaaatag 1649  
 445 GATGTAATAACTCTGAACAGAAGAAGTTTAACTTAGAGTAGCCCTAAATAG 386

1650 gatggtcttataatcgttagtttgagctgtatcgtgaacagagagagctgt 1709  
 385 GATGCTCTTATATATCGCTTAGTTTGGAACGTGTATCGATACAGAGAGAGCTGTT 326

1710 tttaaccctctctcgaagttgttgacctacatggtcctaataatgataa 1769  
 325 TTTTAAACCTCTCTCGAAGTTGTGTGACCTACATGGCTATATGATCTAAATATC 266

1770 tacatgacttaagaagaactagcctctgagagatataagatgctttcattaca 1829  
 265 TACATGACTTAAGAGAAGAACTAGCCTTGAGAGTATATAGTGTTCATTATACNA 206

1830 caaaaatccctgagggacatttgagcagatgaataaacaatttattcagta 1889  
 205 CAAAATCCCTGAGGGACATTTTGAGGATGATATTAACATTTTATTCAGTAACTT 146

1890 tccccctggttaagttaactggttctgtacaaactcattcctataagaata 1949  
 145 TTCCCCCTGTGTAAGTACTATGTTGTGTGTAACACTTCAATATGAAATATTAAGT 86

1950 gaagtggatgaattcactcttattatgttgagtgagcaaatgtcctataa 2009  
 85 GAAGTGGGTGAATCTACTTTTATGTGTGAGTGACCAATGTCTATCAAGATGACAAA 26

2010 taaagtaatatgatccaataa 2032  
 25 TAAAGTTAATGATGATTCACAAA 3

RESULT 42  
 BG538574 878 bp mRNA EST 03-APR-2001  
 LOCUS 602567303F1 NIH\_MGC\_77 Homo sapiens cDNA clone IMAGE:4691712 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG538574  
 VERSION BG538574.1 GI:13530807  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 878)  
 AUTHORS NIH-MGC http://mgs.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)

## COMMENT

Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-femail.nih.gov  
 Tissue Procurement: CLONTECH Laboratories, Inc.  
 cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LLM1511 row: 6 column: 01  
 High quality sequence stop: 615.

## FEATURES

source

1..878  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4691712"  
 /clone\_lib="NIH\_MGC\_77"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site: 1: SfiI (ggccgctggcc); Site 2: SfiI (ggccatattggc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATATGACC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGCGCGCGCATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.9 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."

BASE COUNT 278 a 121 c 200 g 279 t  
 ORIGIN

Query Match 21.8%; Score 443; DB 155; Length 878;  
 Best Local Similarity 99.8%; Pred. No. 2.1e-224;  
 Matches 493; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

764 ggaacgctatgaggaagactctgaacatctcagaagatgctcctaagatcctgtgacg 823  
 1 GGAACTGTTAGGGGAAGACTCTGAACATTCAGAGATGCTTAAAGATCTCTTGACG 60

824 cctcgcaaaagtttgcagatgacatgacatgcttattggtgagatgagtgagag 883  
 61 CTTCTCAAAAGTTTGCAGATGACATGATGATCTTATGATGATGATGATGATGATGAT 120

884 tagtaactgtaagtaacttgcacactccctcattaggaagaagaagactatcctgag 943  
 121 TAGTACTGTAAGTAACTTGTACACCTCCCTCATTTAGGAAGAACAGACTATCTTGAG 180

944 caaaacagaggaagaaacccagaaagctcctataaacttgcatataaataattgga 1003  
 181 CAAAACAGAGGAAGAAACCCAGAAAGCTCCTATTAACCTTCATTAATAGTATTTGAAT 240

1004 attccgtgttttcaacatgtaacttggataatgacgccttgcccttgctgagta 1063  
 241 ATTCCTGTGTTTCAACATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300

1064 taaccttacaatatttggaacatgataccttgatataagatagcatatattagatga 1123  
 301 TACCTCTTCAATATTGTGAACATGATCTCGATATGATGATGATGATGATGATGATGAT 360

1124 caaacagaagatcgaatgagatgaatgtaactcgtgccaagaatlaagaagggttg 1183  
 361 CAAACAGAGATTCGAATGAGATGAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 420

1184 gaattggcgttctgttaaatatatacttattgctcttaagatagatagtaact 1243  
 421 GAATTTGGCTGTTTAAATATATCTTTTAACTGCTTTAAAGTATGATGATGATGATGAT 480

1244 tacattataaaa 1257  
 481 TACATTTTATAAAA 494

RESULT 43  
AA772434 444 bp mRNA EST 31-DEC-1998  
LOCUS a130b06.s1 Soares\_parathyroid\_tumor\_NbHPA Homo sapiens cDNA clone  
DEFINITION 1358483 3', mRNA sequence.  
ACCESSION AA772434  
VERSION AA772434.1 GI:2824217  
KEYWORDS Est.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 444)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo  
, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINL at:  
www-bio.lnl.gov/bdrp/image/image.html  
Insert Length: 808 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 417.

FEATURES  
source  
Location/Qualifiers

1..444  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="1358483"  
/clone\_lib="Soares\_parathyroid\_tumor\_NbHPA"  
/tissue\_type="parathyroid tumor"  
/dev\_stage="adult"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: parathyroid gland; Vector: p773D (Pharmacia  
) with a modified polylinker; Site\_1: Not I; Site\_2: Eco  
RI; 1st strand cDNA was primed with a Not I - oligo(dT)  
primer  
[5'-TGTTACCAATCTGAAGTGGAGCGCGCCACCAATTTTGTGTTTGTGTTT  
TTTGT-3'], double-stranded cDNA was size selected, ligated  
to Eco RI adapters (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of a modified p773  
vector (Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid  
adenomas was kindly provided by Dr. Stephen Marx, National  
Institute of Diabetes and Digestive and Kidney Diseases,  
NIH."

BASE COUNT 148 a 83 c 64 g 149 t  
ORIGIN

Query Match 21.7%; Score 442; DB 11; Length 444;  
Best Local Similarity 100.0%; Pred. No. 6.8e-224;  
Matches 442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1591 atgtaaactctgtaacagaagaagttttaactagtagtaacctaataatag 1650  
Db 444 ATGTAATTAACCTGTAACAGAGAAAGTTTAACTAGTACCTTAATAATAG 385  
Oy 1651 atggtctataataatgctgttggaaactgtatctgagtaacagaagacgtgtt 1710  
Db 384 ATGTCCTTAATATATGCTGTGATTGTAACGTGTATCTGATACACAGACAGCTGTT 325  
Oy 1711 ttaacctcttcgaaagtgttgacctaactgagataatgataaact 1770  
Db 324 TTTAACCCCTTGTGAGATTGTTGACCTACTGCTAATATGATCTAAATACT 265  
Oy 1771 acattgctcaagaagaactagccttgtagatataatagcttttataacac 1830

Db 264 AATTGATCTAGAGAAACTAGCCTTGAGATATATAGCTTTCTATFACAC 205  
Oy 1831 aaaaatccctgagagacatttgagcatataataaacatttatactcagtaactt 1890  
Db 204 AAAAATCCCTGAGGACATTTGAGCATGATATAAACATTTTATTCTACTAATCTT 145  
Oy 1891 tccctctgtaagttactatgttggtaactatctatcatatgataataggg 1950  
Db 144 TCCCTCTGTGATGATCTACTATGCTTTGTGTGACACTTCATCTATGAATATTAAGTG 85  
Oy 1951 aagtgatgtaactctatcttattatgttgagagacatgctatcaagatgacaat 2010  
Db 84 AAGTGGGATGATCTACTTTTATGTTGAGTGGACCAATGCTATCAAGATGACAAAT 25  
Oy 2011 aagttatgattatcccaaa 2032  
Db 24 AAGCTTATATGATGATTCAGAAA 3

RESULT 44  
AA470629 441 bp mRNA EST 13-AUG-1997  
LOCUS AA470629  
DEFINITION ne12b06.s1 NCI\_CGAP\_C03 Homo sapiens cDNA clone IMAGE:880979 3',  
mRNA sequence.  
ACCESSION AA470629  
VERSION AA470629.1 GI:2197938  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 441)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D.  
, Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINL at:  
www-bio.lnl.gov/bdrp/image/image.html  
Insert Length: 1334 Std Error: 0.00  
Seq primer: -41m13 fwd. ET from Amersham  
High quality sequence stop: 323.

FEATURES  
source  
Location/Qualifiers

1..441  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:880979"  
/clone\_lib="NCI\_CGAP\_C03"  
/sex="pooled"  
/tissue\_type="colon"  
/lab\_host="DH10B"  
/note="Vector: p773D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from 12 pooled bulk tumor samples and primed  
with a Not I - oligo(dT) primer. Double-stranded cDNA was  
ligated to Eco RI adapters (Pharmacia), digested with Not  
I and cloned into the Not I and Eco RI sites of the  
modified p773 vector. Library went through one round of  
normalization."

BASE COUNT 148 a 84 c 64 g 145 t  
ORIGIN

Query Match 21.7%; Score 441; DB 7; Length 441;  
Best Local Similarity 100.0%; Pred. No. 2.3e-223;

Matches 441; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1590 gatbaataactcgaagaagaaggttttaacttagtagagcccaaatatg 1649

Db 441 GATGTAAATACCTGGAAGAGAAAGGTTTAACTTAGAGTACCCATAAATATG 382

QY 1650 gatgtctataatcagctagtttggaactgtatcgtgaagaagagcagctgtt 1709

Db 381 GATGTCTATATATATCCCTAGTTTGGAACTGTATGTGATACAGAGACACCTGTT 322

QY 1710 tttaaccccttcgcgaagtttgtagctacatcagggcctaataatgatacacaat 1769

Db 321 TTTTAAACCTCTTCGCAAGTTTGTGACCTCATGAGCTTAATATGATATATAAATATC 262

QY 1770 tacattgataagaagaactagcctgtgagatataatagcttccattataaca 1829

Db 261 TACATTCATCTAAGAGAAACTACCTGTGAGATATATAGATCTTTTCATTATACACA 202

QY 1830 caaaaatccctgagagacatttgaagcagtaataaacaatttattcagtaact 1889

Db 201 CAAAATCCCTGAGGACATTTTGAAGCATGATATATAAATTTTATTTTCAGTACTT 142

QY 1890 ttcccccgtgtaagtactatggttgggtacaaactcattatagataataatg 1949

Db 141 TTCCCTCTGTAAAGTACTATGTTGTGTGACAACTTCAATCTATAGAAATATTAAGTG 82

QY 1950 gaatgggtaacttacttcttattgttgagtgagcaatgtatcaagaagtgacaa 2009

Db 81 GAATGGGTAAATTTACTTTTATGTTGAGTGAGCAATGTCTATCAAGAGTGACAAA 22

QY 2010 taagattaatgatgtccaa 2030

Db 21 TAAAGTTAATGATGATCCAA 1

RESULT 45

LOCUS AA164657 521 bp mRNA EST 09-MAR-1998

DEFINITION z098g1.s1 Stratagene ovarian cancer (#937219) Homo sapiens cDNA

clone IMAGE:594980 3', mRNA sequence.

ACCESSION AA164657

VERSION AA164657.1 GI:1740818

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 521)

Hillier, D., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,

Kritzman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin,

J., Moore, B., Schellenberg, K., Stepien, M., Tan, F., Theisling, B.,

White, Y., Wyllie, T., Waterston, R. and Wilson, R.

WashU-NCI human EST Project

Unpublished (1997)

CONTACT: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

Insert Length: 1427 Std Error: 0.00

Seq primer: 40M13 fwd from Amersham

High quality sequence stop: 338.

FEATURES

source

1. 521

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:594980"

/clone\_lib="Stratagene ovarian cancer (#937219)"

/sex="female"

/dev\_stage="adult, 64 years"

/lab\_host="SOLR (kanamycin resistant)"

/note="Vector: Bluescript SK-; Site\_1: EcoRI; Site\_2: XhoI

; Cloned unidirectionally. Primer: Oligo dT. Papillary

serous carcinoma, isolated from ascites, 64 year old

caucasian. Average insert size: 0.8 kb; Uni-ZAP XR Vector;

-5' adaptor sequence: 5' GAATTCGCGACGAG 3' ~3' adaptor

sequence: 5' CTCGAGTTTATTTTATTTT 3'

BASE COUNT 168 a 99 c 73 g 177 t 4 others

ORIGIN

Query Match 21.6%; Score 440; DB 3; Length 521;

Best Local Similarity 100.0%; Pred. No. 8e-223;

Matches 440; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1593 gtaataactcgaagaagaaggttttaacttagtagagcccaaatatg 1652

Db 450 GTAATATACCTGGAAGAGAAAGGTTTAACTTAGAGTACCCATAAATATGAT 391

QY 1653 gtgctataatcagctagtttggaactgtatcgtgaagaagagcagctgtt 1712

Db 390 GTGCTTATATATCCCTAGTTTGGAACTGTATCTGAGTACAGAGACACCTGTTT 331

QY 1713 taaccccttcgcgaagtttgtagctacatcagggcctaataatgatacacaat 1772

Db 330 TACCCCTCTGTGAAAGTGTGTGACCTCATGAGCTTAATATGATATATAAATATC 271

QY 1773 attgataagaagaactagcctgtgagatataatagcttccattataaca 1832

Db 270 ATTGATCTAAGAGAAACTACCTGTGAGATATATAGATCTTTTCATTATACACA 211

QY 1833 aatccctgagagacatttgaagcagtaataaacaatttattcagtaacttc 1892

Db 210 AATTCCTGAGGACATTTTGAAGCATGATATATAAATTTTATTTTCAGTACTT 151

QY 1893 cccctgtgtaagtactatggttgggtacaaactcattatagataataatg 1949

Db 150 CCCCTGTGTAAGTACTATGTTGTGTGACAACTTCAATCTATAGAAATATTAAGTG 91

QY 1953 gtgagtaacttacttcttattgttgagtgagcaatgtatcaagaagtgacaa 2012

Db 90 GTGGGTGAATTTACTTTTATGTTGAGTGAGCAATGTCTATCAAGATGACAAATTA 31

QY 2013 agttaatgatgtccaa 2032

Db 30 ACTTATGATGATGCCAAA 11

RESULT 46

LOCUS AM958900 624 bp mRNA EST 01-JUN-2000

DEFINITION ES370970 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.

ACCESSION AM958900

VERSION AM958900.1 GI:8148584

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 624)

Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspar, R., Gay, C., Holt,

J., Lee, S., Seed, A., Sharov, V., Lee, N.H., Yeatman, T.J. and

Quackenbush, J.

Assessment of gene expression patterns in a model of colon tumor

metastasis using a 19,200 element cDNA microarray

Unpublished (2000)

CONTACT: John Quackenbush

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 3528

Fax: 301 838 0208

Email: johnq@tigr.org

Plate: 127



```

RESULT 48
AA465651/c 439 bp mRNA EST 13-AUG-1997
LOCUS aa31a05.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:814832 3',
DEFINITION mRNA sequence.
ACCESSION AA465651
VERSION AA465651.1 GI:2191818
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 439)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: 41m3 fwd. ET from Amersham
High quality sequence stop: 413.
Location/Qualifiers
1..439
/organism="Homo sapiens"
/db_xref="GDB:6033470"
/db_xref="taxon:9606"
/clone="IMAGE:814832"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germlinal center B cell"
/lab_host="DH10B"
/molecule="vector: pT73D-Pac (pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germlinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-GTGACCACTGAGAGGAGCGCGCCCTATTTTCTTTTCTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 148 a 83 c 64 g 144 t
ORIGIN

Query Match 21.6%; Score 439; DB 7; Length 439;
Best Local Similarity 100.0%; Pred. No. 2.7e-222;
Matches 439; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1591 atgtaaataactgaaacagaagaaggttttaacttaagaagccctaaataatg 1650
|||||
DB 439 ATGTAAATAACTGAAACAGAAAGAGTTTAACTTAAGAGCCCTAAATAATG 380
|||||
QY 1651 atgtctataataatcgcttagtttggaaactgatactgatacagaagacagctgttt 1710
|||||
DB 379 ATGTCTATAATAATCGCTTAGTTTGAAGTGAATGATGAGTAACAGAGCAGCTGTTT 320
|||||
QY 1711 tttaacctctctgaaagttgttgacctacatgagctaaataatgataactaact 1770
|||||
DB 319 TTAAACCTCTCTTGCAGAGTTGTTGACCTCAATGAGCTAATATGATACATAAATACT 260
|||||
QY 1771 acattgatctaagaagaataagcttgtagatataatagatgtcttcaattacaac 1830
|||||
DB 259 ACATTGATCTAAGAGAACTAGCCTTGCGAGTAATATGATGCTTTTCATTATACACAC 200
|||||
QY 1831 aaaaatccccatgaagacatttgaagcatgaataataaatatttattcaatgaat 1890
|||||
DB 199 AAAATCCCTGAGGACATTTTTCAGCATGAATATAACATTTTATTTCAGTAATCTTT 140
|||||

```

```

QY 1891 tccctctgtgaagtactatggttgggtacaaactcatctctatagaataatg 1950
|||||
DB 139 TCCCTCTGTGTAAGTACTAGTGTGTGTACAACTTCATCTTAAGATATTAAGTG 80
|||||
QY 1951 aagtgtgaatcactcttctatgttggagtgagccaatgtctatcaagaagtacaat 2010
|||||
DB 79 AAGTGGTGAATTCATCTTTTAAAGTTTGGATGGAGCAAGTCTCATCAAGAGTCAAAAT 20
|||||
QY 2011 aaagtaaatgataatcca 2029
|||||
DB 19 AAAGTAAATGATGATTCCA 1
|||||

RESULT 49
BF432282/c 563 bp mRNA EST 30-MAR-2001
LOCUS nab88g04.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:3274950
DEFINITION 3', mRNA sequence.
ACCESSION BF432282
VERSION BF432282.1 GI:11444396
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 563)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL, send email to:
info@image.llnl.gov
Seq primer: -400P from Glibco
High quality sequence stop: 470.
Location/Qualifiers
1..563
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3274950"
/clone_lib="NCI_CGAP_Brn23"
/tissue_type="gliblastoma (pooled)"
/lab_host="DH10B"
/molecule="organ: brain; Vector: pT73D-Pac (pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'-
TGTATCCATCTGTAAGTGGAGCGCGCCGATATCTTTTCTTTTCTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI
adaptors (pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 191 a 104 c 80 g 187 t 1 others
ORIGIN

Query Match 21.5%; Score 437; DB 148; Length 563;
Best Local Similarity 99.8%; Pred. No. 3.2e-221;
Matches 557; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1467 atggaataaagtgcagtaattattatgacaactaagaatgacttaactcttacaca 1526
|||||
DB 563 ATGGAATAAAGTGCAGTAATTATTATGACAACTTACGAATGCTTAATCTTTTACACA 504
|||||
QY 1527 gcatagtggaataatcatttggctatgtatatactatgaacaatttgtaaatgtcttaa 1586
|||||

```

```

Db 503 GCATAGGAGAAATCATATTGGGCTATTGTATATGATGACATTTGTAATGCTTAA 444
QY 1587 ttgtgttaataactcttgaaacaaagaaaggcttttaacttagtagccctaaat 1646
Db 443 TTTGATGTAATTAACCTGTAACAAAGAAAGGTTTAACTTGAAGTACCCCTAAAT 384
QY 1647 atgagtgcttataatagcttagtttgaacgtlaactgaagacagagacagct 1706
Db 383 ATGGATGCTTATATATACGCTAGTTTGAACGTATCTAGCTAACAGAGACAGCT 324
QY 1707 gtttttaaccctctctgaagttgttgacttaacagggttaataacttaaa 1766
Db 323 GTTTTAAACCTCTTCTGCAAGTTTGACCTACATGCGCTAATATGATGACTAAAA 264
QY 1767 tactacattgacttaagaagaactagccttgytggatataatagatgctttcattac 1826
Db 263 TACTACATTTGATCTAAGAAAGAACTAGCCTTGAGATATATAGTGTTCATTATAC 204
QY 1827 acacaaaatccctggaggacatttggagcatgataataaacattttatttcagtaa 1886
Db 203 ACACAAAATTCCTGAGGACATTTTGAGCATGAAATATAAACATTTTATTTCAGTAA 144
QY 1887 ctttc-ccccgtgtaagttactatggttgytgcacacttactataagaatata 1945
Db 143 CTTTTCNCCCTGTGTAAGTACTATGTTGGTGGTACAACTTCTATGATATATTA 84
QY 1946 agtgaagtgaggaaacttacttatttatttgaagtgagcaactgactcaagaagta 2005
Db 83 ACTGGAAGTGGGGAATTTTACTTTTATGTTGAGTGGACCAAGTCTATCAAGATGA 24
QY 2006 caaataagtaagtgatg 2023
Db 23 CAATAAAGTTAATGATG 6

RESULT 50
AA164656 487 bp mRNA EST 09-MAR-1998
LOCUS z098911.r1 Stratagene ovarian cancer (#937219) Homo sapiens cDNA
DEFINITION clone IMAGE:594980 5', mRNA sequence.
ACCESSION AA164656
VERSION AA164656.1 GI:1740817
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 487)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Kizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wyllie,T., Waterston,R. and Wilson,R.
Mashu-NCI human EST Project
Unpublished (1997)
CONTACT: Wilson RK
JOURNAL Washington University School of Medicine
COMMENT 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1427 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 391.
Location/Qualifiers
1..487
FEATURES
Source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:594980"
/clone_lib="Stratagene ovarian cancer (#937219)"
/sex="Female"

```

```

/dev_stage="adult, 64 years"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: Bluescript SK-; site_1: EcoRI; site_2: XhoI
; Cloned unidirectionally. Primer: Oligo dT. Papillary
serous carcinoma, isolated from ascites, 64 year old
caucasian. Average insert size: 0.8 kb; Uni-ZAP XR Vector;
-5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor
sequence: 5' CTCGACTTTTCTTTTCTTTTCTTTT 3'"
BASP COUNT 145 a 88 c 111 g 142 t 1 others
ORIGIN

Query Match 21.4%; Score 436; DB 3; Length 487;
Best local similarity 99.8%; Pred. No. 1.1e-220;
Matches 486; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 693 tcatagacatctgccaagagatctctctgattatattcactggagctgacagttt 752
Db 1 TCATAGCATCTAGCCAAAGATCATCTCTGATTTATATTCACCTGGAGCTGCGAGGTTT 60
QY 753 ggaatgaattgggaagcgttctgggaagactctgaacatccagaatgcttcaagat 812
Db 61 GGATGAATTTGGGAACGCTTATGGGGAAGCTCTGACAAATTTCAGAGATGCTTAAAT 120
QY 813 ccttgtagagctctgcaaaagtttgagatgacatgacagcttatagttggaaatgc 872
Db 121 CCTTGTTGACGCTCTGCAAAAGTTTGCAGATGACATGTCATTTATGTTGGGAATGC 180
QY 873 agtggtagttagtcaactgcaagcttgaacacccctccatlaggaagacaagac 932
Db 181 AGTGTAGAGTGTAGTACACTGCAAGTCATTTGACACCTCCCTATTAGGAACAGAGAC 240
QY 933 tatccttgagcaaaacaaaggaagaaacccagaagctccataacttgacataagta 992
Db 241 TATCCTTGAGGCAAAACAAAGCAAGAACCCAGCAAGTCCCTATACCTTGCAATAAAGTA 300
QY 993 taatttgaatattcgcgtgttttcaacatggttacttgataatgatcgccttgcctt 1052
Db 301 TAATTTGAATATTCCTGCTGTTTCACATGATGACTTTGATATATGATGCTGCTTGCCTT 360
QY 1053 ggcgtgatatcactcttcaatatttggacaatgagatcctgagatagataagacatc 1112
Db 361 GCGTGTATATACCTCTTACAAATTTTGGAACTGATGCTGATGATGATGATGATCAT 420
QY 1113 ttatagatgacaacacagaagatlcgaatggaatggaatggttaccctgtgccagaattaga 1172
Db 421 TTATAGATACCAAAACCAAGAGATGCAATGATGATGATGATGATGATGATGATGATGAT 480
QY 1173 aaagggg 1179
Db 481 AAAGGGG 487

Search completed: August 18, 2001, 22:54:07
Job time: 12169 sec

```



Mon Aug 20 10:21:43 2001

us-09-284-320-56.0110.rst

---

Page 41



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2001, 00:19:56 ; Search time 117.81 Seconds

(without alignments)  
3266.866 Million cell updates/sec

Title: US-09-284-320-56

Perfect score: 2033

Sequence: 1 gagtcgagcgcgtacctc.....gttaatgatgccaaac 2033

Scoring table: OLIGO\_MNC  
Gapop 60.0 , Gapext 60.0

Searched: 324599 seqs, 9465562 residues

Word size: 10

Total number of hits satisfying chosen parameters: 110463

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :  
1: /cgnl\_7/ptodata/1/ina/5A.COMB.seq.\*  
2: /cgnl\_7/ptodata/1/ina/5B.COMB.seq.\*  
3: /cgnl\_7/ptodata/1/ina/6A.COMB.seq.\*  
4: /cgnl\_7/ptodata/1/ina/6B.COMB.seq.\*  
5: /cgnl\_7/ptodata/1/ina/PCRTS.COMB.seq.\*  
6: /cgnl\_7/ptodata/1/ina/backfiles1.seq.\*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	227	11.2	343	4	US-08-905-223-169
2	19	0.9	174	6	Sequence 169, App
3	19	0.9	420	2	Patent No. 519031-2
4	19	0.9	500	6	Sequence 20, Appl
5	19	0.9	1416	2	Patent No. 519031-1
6	18	0.9	643	4	Sequence 27, Appl
7	18	0.9	877	3	Sequence 1022, Ap
8	18	0.9	1843	5	Sequence 1, Appli
9	18	0.9	1843	5	Sequence 1, Appli
10	18	0.9	2083	2	PCT-US95-03323A-1
11	17	0.8	155	1	Sequence 1, Appli
12	17	0.8	156	1	Sequence 8, Appli
13	17	0.8	163	1	Sequence 9, Appli
14	17	0.8	210	1	Sequence 11, Appl
15	17	0.8	211	1	Sequence 13, Appl
16	17	0.8	319	1	Sequence 12, Appl
17	17	0.8	510	4	Sequence 14, Appl
18	17	0.8	527	4	Sequence 79, Appl
19	17	0.8	822	4	Sequence 42, Appl
20	17	0.8	1104	5	Sequence 19, Appl
21	17	0.8	1241	4	Sequence 6, Appli
22	17	0.8	1241	4	Sequence 6, Appli
23	17	0.8	1596	1	Sequence 3, Appli
24	17	0.8	1596	2	Sequence 3, Appli
25	17	0.8	1596	2	Sequence 3, Appli
26	17	0.8	1596	4	Sequence 3, Appli
27	17	0.8	1633	3	Sequence 4, Appli

C 28	17	0.8	1721	3	US-08-857-213-2	Sequence 2, Appli
C 29	17	0.8	2307	3	US-08-942-008-1	Sequence 1, Appli
C 30	17	0.8	2986	3	US-09-062-416-1	Sequence 1, Appli
C 31	17	0.8	3016	2	US-08-344-155C-97	Sequence 97, Appl
C 32	17	0.8	3024	6	5284931-1	Patent No. 5284931
C 33	17	0.8	3435	1	US-08-366-577-1	Sequence 1, Appli
C 34	17	0.8	3435	5	PCT-US96-00005-1	Sequence 1, Appli
C 35	17	0.8	4624	5	US-08-696-834-5	Sequence 5, Appli
C 36	17	0.8	4624	4	US-09-077-354B-3	Sequence 3, Appli
C 37	17	0.8	10380	2	US-08-670-175-2	Sequence 2, Appli
C 38	16	0.8	52	2	US-08-220-151-41	Sequence 41, Appl
C 39	16	0.8	66	1	US-08-413-118-41	Sequence 41, Appl
C 40	16	0.8	66	1	US-08-224-657-18	Sequence 18, Appl
C 41	16	0.8	66	1	US-08-257-073-85	Sequence 85, Appl
C 42	16	0.8	66	2	US-08-184-009-18	Sequence 18, Appl
C 43	16	0.8	66	2	US-08-486-969-18	Sequence 18, Appl
C 44	16	0.8	66	2	US-08-417-210A-18	Sequence 18, Appl
C 45	16	0.8	66	2	US-08-458-336-18	Sequence 18, Appl
C 46	16	0.8	66	2	US-08-471-025-18	Sequence 18, Appl
C 47	16	0.8	66	3	US-08-473-446-41	Sequence 41, Appl
C 48	16	0.8	66	5	PCT-US96-00547-18	Sequence 4, Appli
C 49	16	0.8	148	1	US-08-270-985-4	Sequence 4, Appli
C 50	16	0.8	148	3	US-08-478-208-5	Sequence 5, Appli
C 51	16	0.8	250	1	US-08-464-164-5	Sequence 5, Appli
C 52	16	0.8	250	2	US-08-338-057-5	Sequence 5, Appli
C 53	16	0.8	250	2	US-08-668-416-5	Sequence 5, Appli
C 54	16	0.8	318	1	US-08-702-344-17	Sequence 6, Appli
C 55	16	0.8	399	3	US-08-713-569-6	Sequence 6, Appli
C 56	16	0.8	660	4	US-08-990-791-1	Sequence 1, Appli
C 57	16	0.8	917	4	US-08-942-012B-1	Sequence 4, Appli
C 58	16	0.8	924	1	US-07-593-657-4	Sequence 4, Appli
C 59	16	0.8	988	1	US-08-684-862-10	Sequence 10, Appl
C 60	16	0.8	996	1	US-08-671-525B-7	Sequence 7, Appli
C 61	16	0.8	996	1	US-08-672-109B-7	Sequence 7, Appli
C 62	16	0.8	996	1	US-08-842-045-7	Sequence 7, Appli
C 63	16	0.8	996	2	US-08-842-238-7	Sequence 7, Appli
C 64	16	0.8	996	3	US-08-629-335B-7	Sequence 7, Appli
C 65	16	0.8	1239	1	US-08-477-254A-3	Sequence 3, Appli
C 66	16	0.8	1239	2	US-08-472-576B-3	Sequence 3, Appli
C 67	16	0.8	1239	2	US-08-428-734B-3	Sequence 3, Appli
C 68	16	0.8	1354	1	US-08-514-014-9	Sequence 9, Appli
C 69	16	0.8	1354	2	US-08-833-823-9	Sequence 9, Appli
C 70	16	0.8	1361	2	US-08-387-892-9	Sequence 9, Appli
C 71	16	0.8	1361	2	US-08-387-858A-9	Sequence 9, Appli
C 72	16	0.8	1361	4	US-09-294-384B-9	Sequence 9, Appli
C 73	16	0.8	1520	1	US-08-470-187-7	Sequence 7, Appli
C 74	16	0.8	1520	1	US-08-318-905-7	Sequence 7, Appli
C 75	16	0.8	1520	1	US-08-483-232-7	Sequence 7, Appli
C 76	16	0.8	1520	1	US-08-483-140-7	Sequence 7, Appli
C 77	16	0.8	1520	2	US-08-485-938A-7	Sequence 7, Appli
C 78	16	0.8	1520	2	US-08-910-041-7	Sequence 7, Appli
C 79	16	0.8	1520	3	US-09-328-474-7	Sequence 7, Appli
C 80	16	0.8	1520	3	US-09-100-546-7	Sequence 7, Appli
C 81	16	0.8	1520	4	US-09-010-715-7	Sequence 7, Appli
C 82	16	0.8	1520	4	US-09-577-758-7	Sequence 7, Appli
C 83	16	0.8	1587	2	US-08-477-254A-6	Sequence 6, Appli
C 84	16	0.8	1587	2	US-08-472-576B-6	Sequence 6, Appli
C 85	16	0.8	1587	2	US-08-428-734B-6	Sequence 6, Appli
C 86	16	0.8	1641	3	US-08-888-998-1	Sequence 1, Appli
C 87	16	0.8	1641	4	US-09-362-633-1	Sequence 1, Appli
C 88	16	0.8	1649	1	US-08-477-254A-1	Sequence 1, Appli
C 89	16	0.8	1649	2	US-08-472-576B-1	Sequence 1, Appli
C 90	16	0.8	1649	2	US-08-428-734B-1	Sequence 1, Appli
C 91	16	0.8	1671	2	US-08-662-560-1	Sequence 1, Appli
C 92	16	0.8	1671	2	US-08-780-249A-5	Sequence 5, Appli
C 93	16	0.8	1671	2	US-08-706-281A-15	Sequence 15, Appl
C 94	16	0.8	1977	2	US-08-825-558-3	Sequence 3, Appli
C 95	16	0.8	1984	4	US-08-822-028-16	Sequence 16, Appl
C 96	16	0.8	1984	4	US-08-479-285-16	Sequence 16, Appl
C 97	16	0.8	2019	2	US-08-455-073A-5	Sequence 5, Appli
C 98	16	0.8	2042	2	US-08-933-821-16	Sequence 16, Appl
C 99	16	0.8	2042	3	US-08-934-494-5	Sequence 5, Appli
C 100	16	0.8	2042	3	US-08-960-507-16	Sequence 16, Appl

101	16	0.8	2042	3	US-09-143-068-5	Sequence 5, Appl	c 174	15	0.7	483	1	US-08-592-126-118	Sequence 118, App
c 102	16	0.8	2280	1	US-08-415-818-5	Sequence 5, Appl	c 175	15	0.7	484	3	US-08-899-786-19	Sequence 19, Appl
c 103	16	0.8	2280	2	US-08-894-236-5	Sequence 5, Appl	c 176	15	0.7	485	1	US-08-249-671A-6	Sequence 6, Appl
c 104	16	0.8	2280	1	US-09-0144-5	Sequence 5, Appl	c 177	15	0.7	522	4	US-09-277-565-18	Sequence 18, Appl
105	16	0.8	2369	1	US-07-797-556-1	Sequence 1, Appl	c 178	15	0.7	540	6	US-03-277-565-18	Patent No. 5310729
106	16	0.8	2369	1	US-08-308-881-1	Sequence 1, Appl	c 179	15	0.7	564	4	US-08-998-416-843	Sequence 843, App
107	16	0.8	2369	2	US-09-058-263-1	Sequence 1, Appl	c 180	15	0.7	648	2	US-08-659-251-11	Sequence 11, Appl
108	16	0.8	2369	2	US-09-059-099-1	Sequence 1, Appl	c 181	15	0.7	648	5	US-09-256-490-11	Sequence 11, Appl
109	16	0.8	2369	3	US-09-058-264-1	Sequence 1, Appl	c 182	15	0.7	648	5	PCT-US96-11445-11	Sequence 57, Appl
110	16	0.8	2369	5	PCT-US95-06530-1	Sequence 1, Appl	c 183	15	0.7	651	2	US-08-912-129A-57	Sequence 1, Appl
111	16	0.8	2372	1	US-07-903-103-1	Sequence 1, Appl	c 184	15	0.7	704	3	US-09-009-136-1	Sequence 3, Appl
112	16	0.8	2372	1	US-08-044-619A-1	Sequence 1, Appl	c 185	15	0.7	711	1	US-08-266-570A-3	Sequence 1, Appl
113	16	0.8	2372	1	US-08-283-911-1	Sequence 1, Appl	c 186	15	0.7	720	6	US-08-691-814B-116	Patent No. 5310729
114	16	0.8	2372	1	US-08-245-500A-2	Sequence 2, Appl	c 187	15	0.7	737	2	US-08-631-814B-116	Sequence 116, App
115	16	0.8	2372	1	US-08-390-546C-2	Sequence 2, Appl	c 188	15	0.7	741	2	US-08-912-129A-47	Sequence 47, Appl
116	16	0.8	2372	1	US-08-390-479A-2	Sequence 2, Appl	c 189	15	0.7	757	4	US-08-998-416-454	Sequence 454, App
117	16	0.8	2372	1	US-08-557-593-2	Sequence 2, Appl	c 190	15	0.7	791	2	US-07-635-196A-4	Sequence 4, App
118	16	0.8	2372	1	US-08-390-516C-2	Sequence 2, Appl	c 191	15	0.7	799	1	US-08-173-510B-100	Sequence 100, App
119	16	0.8	2372	1	US-08-390-517A-2	Sequence 2, Appl	c 192	15	0.7	799	1	US-08-458-218-96	Sequence 98, Appl
120	16	0.8	2372	1	US-08-390-515A-2	Sequence 2, Appl	c 193	15	0.7	799	2	US-08-450-497-100	Sequence 100, App
121	16	0.8	2372	2	US-08-801-718-2	Sequence 2, Appl	c 194	15	0.7	831	4	US-08-998-416-513	Sequence 513, App
122	16	0.8	2372	2	US-09-073-567-1	Sequence 2, Appl	c 195	15	0.7	845	1	US-08-266-451B-1	Patent No. 532081
123	16	0.8	2372	4	US-09-280-805-1	Sequence 1, Appl	c 196	15	0.7	845	1	US-08-459-019A-3	Sequence 3, Appl
124	16	0.8	2372	4	US-09-048-810-1	Sequence 1, Appl	c 197	15	0.7	855	2	US-08-748-725-1	Sequence 1, Appl
c 125	16	0.8	2433	4	US-08-945-056-3	Sequence 1, Appl	c 198	15	0.7	855	4	US-09-060-726A-1	Sequence 1, Appl
126	16	0.8	2557	2	US-09-044-506A-1	Sequence 3, Appl	c 199	15	0.7	856	4	US-09-060-726A-3	Sequence 3, Appl
127	16	0.8	2557	2	US-08-478-208-31	Sequence 31, Appl	c 200	15	0.7	869	1	US-08-249-671A-7	Sequence 7, Appl
128	16	0.8	2557	6	5264554-1	Patent No. 5264554	c 201	15	0.7	879	1	US-08-249-671A-10	Sequence 10, Appl
129	16	0.8	2754	2	US-08-825-558-5	Sequence 5, Appl	c 202	15	0.7	940	5	PCT-US94-09450-11	Sequence 11, Appl
130	16	0.8	2814	2	US-08-743-637B-186	Sequence 186, App	c 203	15	0.7	946	3	US-09-188-930-258	Sequence 258, App
c 131	16	0.8	2917	2	US-08-437-607A-3	Sequence 3, Appl	c 204	15	0.7	966	1	US-07-637-250A-8	Sequence 8, Appl
c 132	16	0.8	3085	4	US-08-795-473B-4	Sequence 4, Appl	c 205	15	0.7	986	1	US-08-145-061-8	Sequence 1, Appl
c 133	16	0.8	3384	2	US-08-735-609-9	Sequence 9, Appl	c 206	15	0.7	1016	1	US-08-109-391A-3	Sequence 3, Appl
c 134	16	0.8	3384	2	US-08-735-609-9	Sequence 9, Appl	c 207	15	0.7	1016	2	US-08-458-860A-3	Sequence 3, Appl
c 135	16	0.8	3364	3	US-09-315-372-9	Sequence 9, Appl	c 208	15	0.7	1016	2	US-08-933-750C-90	Sequence 90, Appl
c 136	16	0.8	3364	3	US-09-244-752-9	Sequence 9, Appl	c 209	15	0.7	1016	2	US-08-459-019A-3	Sequence 3, Appl
c 137	16	0.8	3364	3	US-09-245-497-6	Sequence 9, Appl	c 210	15	0.7	1078	3	US-08-933-750C-90	Sequence 90, Appl
c 138	16	0.8	3827	1	US-08-170-294-9	Sequence 9, Appl	c 211	15	0.7	1089	5	PCT-US94-09450-7	Sequence 7, Appl
c 139	16	0.8	3827	2	US-08-664-855-6	Sequence 6, Appl	c 212	15	0.7	1089	2	US-09-076-193-2	Sequence 2, Appl
c 140	16	0.8	3827	2	US-08-718-751-1	Sequence 6, Appl	c 213	15	0.7	1095	1	US-08-423-564-3	Sequence 3, Appl
c 141	16	0.8	3827	2	US-08-049-289-6	Sequence 6, Appl	c 214	15	0.7	1107	2	US-08-933-750C-77	Sequence 77, Appl
c 142	16	0.8	4000	2	US-08-861-464-5	Sequence 1, Appl	c 215	15	0.7	1121	3	US-09-234-613-77	Sequence 3, Appl
c 143	16	0.8	4000	2	US-08-396-001-5	Sequence 5, Appl	c 216	15	0.7	1125	2	US-08-266-570A-5	Sequence 5, Appl
c 144	16	0.8	4000	4	US-09-323-433A-5	Sequence 5, Appl	c 217	15	0.7	1203	2	US-08-961-106-3	Sequence 3, Appl
c 145	16	0.8	4566	1	US-08-484-101B-7	Sequence 37, Appl	c 218	15	0.7	1267	3	US-08-961-083-115	Sequence 115, App
c 146	16	0.8	6000	4	US-09-006-636-9	Sequence 9, Appl	c 219	15	0.7	1271	1	US-08-474-542A-141	Sequence 141, App
c 147	16	0.8	6000	4	US-09-006-636-9	Sequence 9, Appl	c 220	15	0.7	1271	1	US-08-457-648-141	Sequence 141, App
c 148	16	0.8	6763	2	US-08-756-506-23	Sequence 23, Appl	c 221	15	0.7	1273	3	US-09-116-622-1	Sequence 1, Appl
c 149	16	0.8	8067	1	US-08-117-362-8	Sequence 8, Appl	c 222	15	0.7	1273	4	US-09-219-277-1	Sequence 1, Appl
c 150	16	0.8	8067	1	US-08-486-924-8	Sequence 8, Appl	c 223	15	0.7	1287	1	US-08-064-121-3	Sequence 1, Appl
c 151	16	0.8	10607	2	US-08-639-857-32	Sequence 32, Appl	c 224	15	0.7	1287	1	US-08-423-564-4	Sequence 4, Appl
c 152	16	0.8	10607	1	US-08-078-090-3	Sequence 3, Appl	c 225	15	0.7	1287	3	US-08-475-975-3	Sequence 3, Appl
c 153	16	0.8	10807	2	US-08-206-176-7	Sequence 7, Appl	c 226	15	0.7	1287	3	US-08-475-975-3	Sequence 3, Appl
c 154	16	0.8	10807	2	US-08-756-506-5	Sequence 5, Appl	c 227	15	0.7	1287	3	US-08-475-975-3	Sequence 3, Appl
c 155	15	0.7	15	3	US-08-292-620A-355	Sequence 355, App	c 228	15	0.7	1287	3	US-08-475-975-3	Sequence 3, Appl
c 156	15	0.7	15	3	US-09-071-845-355	Sequence 355, App	c 229	15	0.7	1287	3	US-08-475-975-3	Sequence 3, Appl
c 157	15	0.7	15	3	US-08-912-128A-35	Sequence 35, App	c 230	15	0.7	1287	3	US-08-475-975-3	Sequence 3, Appl
c 158	15	0.7	20	4	US-09-280-805-96	Sequence 96, Appl	c 231	15	0.7	1287	3	US-08-475-975-3	Sequence 3, Appl
c 159	15	0.7	24	6	5182196-19	Patent No. 5182196	c 232	15	0.7	1287	3	US-08-475-975-3	Sequence 3, Appl
c 160	15	0.7	27	1	US-08-538-875-12	Sequence 12, Appl	c 233	15	0.7	1287	3	US-08-475-975-3	Sequence 3, Appl
c 161	15	0.7	43	1	US-08-356-405-10	Sequence 10, Appl	c 234	15	0.7	1287	3	US-08-475-975-3	Sequence 3, Appl
c 162	15	0.7	101	2	US-08-912-129A-8	Sequence 12, Appl	c 235	15	0.7	1287	3	US-08-475-975-3	Sequence 3, Appl
c 163	15	0.7	111	4	US-09-277-565-22	Sequence 22, Appl	c 236	15	0.7	1287	3	US-08-475-975-3	Sequence 3, Appl
c 164	15	0.7	260	4	US-08-134-198E-11	Sequence 11, Appl	c 237	15	0.7	1287	3	US-08-475-975-3	Sequence 3, Appl
c 165	15	0.7	260	4	5514558-9	Sequence 11, Appl	c 238	15	0.7	1287	3	US-08-475-975-3	Sequence 3, Appl
c 166	15	0.7	306	4	US-09-122-400B-9	Sequence 9, Appl	c 239	15	0.7	1287	3	US-08-475-975-3	Sequence 3, Appl
c 167	15	0.7	348	2	US-08-623-906A-14	Sequence 14, Appl	c 240	15	0.7	1287	3	US-08-475-975-3	Sequence 3, Appl
c 168	15	0.7	423	1	US-08-470-179-38	Sequence 38, Appl	c 241	15	0.7	1287	3	US-08-475-975-3	Sequence 3, Appl
c 169	15	0.7	432	1	US-07-885-970A-11	Sequence 11, Appl	c 242	15	0.7	1287	3	US-08-475-975-3	Sequence 3, Appl
c 170	15	0.7	432	1	US-08-298-687A-11	Sequence 11, Appl	c 243	15	0.7	1287	3	US-08-475-975-3	Sequence 3, Appl
c 171	15	0.7	432	1	US-08-530-797-10	Sequence 10, Appl	c 244	15	0.7	1287	3	US-08-475-975-3	Sequence 3, Appl
c 172	15	0.7	432	1	US-08-298-829-11	Sequence 11, Appl	c 245	15	0.7	1287	3	US-08-475-975-3	Sequence 3, Appl
c 173	15	0.7	432	2	US-08-787-335-10	Sequence 10, Appl	c 246	15	0.7	1287	3	US-08-475-975-3	Sequence 3, Appl

247	15	0.7	1526	2	US-08-999-811-3	Sequence 3, Appl	320	15	0.7	2116	2	US-08-663-566A-18	Sequence 18, Appl
248	15	0.7	1526	3	US-09-042-105-3	Sequence 3, Appl	321	15	0.7	2116	2	US-08-023-610-18	Sequence 18, Appl
C 249	15	0.7	1530	3	US-08-704-931-3	Sequence 3, Appl	322	15	0.7	2116	2	US-08-288-065A-18	Sequence 18, Appl
C 250	15	0.7	1544	4	US-08-991-677-9	Sequence 9, Appl	323	15	0.7	2116	2	US-08-362-240A-18	Sequence 18, Appl
C 251	15	0.7	1586	1	US-08-461-244-1	Sequence 1, Appl	324	15	0.7	2116	5	PCT-US95-10245-18	Sequence 7, Appl
C 252	15	0.7	1650	1	US-08-999-561-5	Sequence 5, Appl	325	15	0.7	2168	1	US-08-423-564-7	Sequence 7, Appl
C 253	15	0.7	1674	2	US-08-999-811-1	Sequence 1, Appl	326	15	0.7	2230	1	US-08-217-327-5	Sequence 1, Appl
C 254	15	0.7	1674	2	US-09-042-105-1	Sequence 1, Appl	327	15	0.7	2255	1	US-08-423-564-1	Sequence 1, Appl
C 255	15	0.7	1674	6	US-07-945-283-3	Sequence 3, Appl	328	15	0.7	2296	2	US-08-951-648-1	Sequence 1, Appl
C 256	15	0.7	1683	1	US-08-747-221B-33	Sequence 33, Appl	C 329	15	0.7	2298	2	US-09-174-437-1	Sequence 6, Appl
C 257	15	0.7	1710	3	US-07-687-466B-1	Sequence 1, Appl	C 330	15	0.7	2325	2	US-08-982-232-6	Sequence 13, Appl
C 258	15	0.7	1718	1	US-08-434-702-1	Sequence 1, Appl	332	15	0.7	2329	2	US-08-982-232-13	Sequence 5, Appl
C 259	15	0.7	1718	1	US-08-271-883-1	Sequence 1, Appl	333	15	0.7	2359	4	US-09-233-813-5	Sequence 5, Appl
C 260	15	0.7	1718	6	5168064-3	Sequence 3, Appl	334	15	0.7	2359	4	PCT-US95-09470-5	Sequence 5, Appl
C 261	15	0.7	1721	1	US-07-688-352C-13	Sequence 13, Appl	335	15	0.7	2433	1	US-08-248-021A-1	Sequence 2, Appl
C 262	15	0.7	1721	2	US-08-474-379C-13	Sequence 13, Appl	336	15	0.7	2433	2	US-08-745-934-2	Sequence 2, Appl
C 263	15	0.7	1721	3	US-09-146-249A-13	Sequence 13, Appl	C 337	15	0.7	2443	2	US-08-463-081B-5	Sequence 5, Appl
C 264	15	0.7	1721	3	US-08-206-188B-13	Sequence 13, Appl	338	15	0.7	2450	2	US-08-461-379A-5	Sequence 5, Appl
C 265	15	0.7	1721	3	PCT-US91-02714-13	Sequence 13, Appl	339	15	0.7	2450	2	US-08-462-390B-5	Sequence 5, Appl
C 266	15	0.7	1785	3	US-08-747-221B-34	Sequence 34, Appl	340	15	0.7	2450	3	US-08-463-074B-5	Sequence 5, Appl
C 267	15	0.7	1785	3	US-08-747-221B-35	Sequence 35, Appl	341	15	0.7	2450	3	US-08-465-585C-5	Sequence 5, Appl
C 268	15	0.7	1812	2	US-08-735-041A-3	Sequence 3, Appl	342	15	0.7	2450	3	US-08-652-446-5	Sequence 5, Appl
C 269	15	0.7	1812	3	US-09-190-476B-3	Sequence 3, Appl	343	15	0.7	2489	3	US-09-141-047-7	Sequence 7, Appl
C 270	15	0.7	1812	3	US-09-190-889A-3	Sequence 3, Appl	C 344	15	0.7	2538	3	US-08-899-437-1	Sequence 1, Appl
C 271	15	0.7	1812	4	US-09-190-889A-3	Sequence 3, Appl	345	15	0.7	2538	4	US-09-126-121-1	Sequence 1, Appl
C 272	15	0.7	1812	4	US-09-190-938B-3	Sequence 3, Appl	346	15	0.7	2623	2	US-08-818-514-1	Sequence 1, Appl
C 273	15	0.7	1831	6	5215881-1	Sequence 53, Appl	C 347	15	0.7	2623	2	US-08-818-514-2	Sequence 2, Appl
C 274	15	0.7	1831	6	5215881-3	Sequence 53, Appl	348	15	0.7	2634	3	US-09-115-934A-1	Sequence 2, Appl
C 275	15	0.7	1860	2	US-08-759-581B-19	Sequence 19, Appl	C 349	15	0.7	2634	3	US-09-115-934A-2	Sequence 2, Appl
C 276	15	0.7	1864	2	US-08-458-120-1	Sequence 1, Appl	350	15	0.7	2634	3	US-09-115-934A-2	Sequence 2, Appl
C 277	15	0.7	1882	1	US-08-458-120-1	Sequence 1, Appl	C 351	15	0.7	2634	3	US-07-977-434-11	Sequence 11, Appl
C 278	15	0.7	1882	2	US-08-867-970-1	Sequence 1, Appl	C 352	15	0.7	2679	1	US-08-458-819-11	Sequence 11, Appl
C 279	15	0.7	1882	2	US-08-867-970-1	Sequence 1, Appl	C 353	15	0.7	2679	1	PCT-US91-07035-11	Sequence 11, Appl
C 280	15	0.7	1882	4	US-09-326-217-1	Sequence 1, Appl	C 354	15	0.7	2679	5	US-08-909-965C-1	Sequence 1, Appl
C 281	15	0.7	1882	4	US-09-326-217-1	Sequence 1, Appl	355	15	0.7	2688	2	US-08-465-995A-1	Sequence 1, Appl
C 282	15	0.7	1882	4	US-09-326-217-1	Sequence 1, Appl	356	15	0.7	2694	2	US-08-465-995A-3	Sequence 1, Appl
C 283	15	0.7	1901	5	PCT-US93-05000-12	Sequence 32, Appl	357	15	0.7	2694	2	US-08-465-994C-1	Sequence 1, Appl
C 284	15	0.7	1931	3	US-09-019-942-2	Sequence 77, Appl	358	15	0.7	2694	2	US-08-465-994C-3	Sequence 3, Appl
C 285	15	0.7	1945	3	US-08-714-918-77	Sequence 77, Appl	359	15	0.7	2694	2	US-08-966-145-1	Sequence 3, Appl
C 286	15	0.7	1945	4	US-09-265-315-77	Sequence 77, Appl	360	15	0.7	2694	2	US-08-966-145-3	Sequence 3, Appl
C 287	15	0.7	1945	4	US-09-265-315-77	Sequence 77, Appl	361	15	0.7	2694	2	US-08-101-593-1	Sequence 1, Appl
C 288	15	0.7	1945	4	US-09-265-315-77	Sequence 77, Appl	362	15	0.7	2760	1	US-08-101-593-3	Sequence 3, Appl
C 289	15	0.7	1948	1	US-07-849-438-1	Sequence 1, Appl	363	15	0.7	2760	1	US-08-101-593-3	Sequence 3, Appl
C 290	15	0.7	1948	1	US-07-866-560-3	Sequence 3, Appl	364	15	0.7	2801	3	US-08-747-221B-30	Sequence 30, Appl
C 291	15	0.7	2012	1	US-08-077-673-3	Sequence 3, Appl	365	15	0.7	2815	1	US-08-230-491A-1	Sequence 1, Appl
C 292	15	0.7	2012	1	US-08-478-992-3	Sequence 3, Appl	C 366	15	0.7	2815	1	US-08-619-280A-1	Sequence 1, Appl
C 293	15	0.7	2012	3	US-09-105-298-3	Sequence 3, Appl	367	15	0.7	2815	2	US-09-018-628-17	Sequence 17, Appl
C 294	15	0.7	2012	3	US-08-464-517-32	Sequence 32, Appl	368	15	0.7	2815	2	US-08-619-280A-1	Sequence 17, Appl
C 295	15	0.7	2022	2	US-08-246-361A-32	Sequence 32, Appl	369	15	0.7	2907	3	US-08-605-150A-11	Sequence 11, Appl
C 296	15	0.7	2022	2	US-08-246-361A-32	Sequence 32, Appl	370	15	0.7	2907	3	US-08-929-329-2	Sequence 2, Appl
C 297	15	0.7	2022	2	US-08-463-772-32	Sequence 32, Appl	C 371	15	0.7	3126	3	US-09-101-886B-1	Sequence 1, Appl
C 298	15	0.7	2029	6	5457090-1	Sequence 3, Appl	C 372	15	0.7	3147	4	US-08-413-135-1	Sequence 5, Appl
C 299	15	0.7	2029	6	5457090-1	Sequence 3, Appl	C 373	15	0.7	3147	4	US-08-951-648-5	Sequence 5, Appl
C 300	15	0.7	2029	6	5457090-1	Sequence 3, Appl	C 374	15	0.7	3195	2	US-09-174-437-5	Sequence 5, Appl
C 301	15	0.7	2029	6	5457090-1	Sequence 3, Appl	C 375	15	0.7	3195	2	US-08-335-844A-14	Sequence 14, Appl
C 302	15	0.7	2029	6	5457090-1	Sequence 3, Appl	C 376	15	0.7	3195	2	US-08-335-844A-15	Sequence 15, Appl
C 303	15	0.7	2029	6	5457090-1	Sequence 3, Appl	C 377	15	0.7	3195	2	US-08-335-844A-15	Sequence 15, Appl
C 304	15	0.7	2029	6	5457090-1	Sequence 3, Appl	C 378	15	0.7	3195	2	US-08-335-844A-15	Sequence 15, Appl
C 305	15	0.7	2029	6	5457090-1	Sequence 3, Appl	C 379	15	0.7	3195	2	US-08-335-844A-15	Sequence 15, Appl
C 306	15	0.7	2029	6	5457090-1	Sequence 3, Appl	C 380	15	0.7	3195	2	US-08-335-844A-15	Sequence 15, Appl
C 307	15	0.7	2029	6	5457090-1	Sequence 3, Appl	C 381	15	0.7	3195	2	US-08-335-844A-15	Sequence 15, Appl
C 308	15	0.7	2029	6	5457090-1	Sequence 3, Appl	C 382	15	0.7	3195	2	US-08-335-844A-15	Sequence 15, Appl
C 309	15	0.7	2029	6	5457090-1	Sequence 3, Appl	C 383	15	0.7	3195	2	US-08-335-844A-15	Sequence 15, Appl
C 310	15	0.7	2029	6	5457090-1	Sequence 3, Appl	C 384	15	0.7	3195	2	US-08-335-844A-15	Sequence 15, Appl
C 311	15	0.7	2029	6	5457090-1	Sequence 3, Appl	C 385	15	0.7	3195	2	US-08-335-844A-15	Sequence 15, Appl
C 312	15	0.7	2029	6	5457090-1	Sequence 3, Appl	C 386	15	0.7	3195	2	US-08-335-844A-15	Sequence 15, Appl
C 313	15	0.7	2029	6	5457090-1	Sequence 3, Appl	C 387	15	0.7	3195	2	US-08-335-844A-15	Sequence 15, Appl
C 314	15	0.7	2029	6	5457090-1	Sequence 3, Appl	C 388	15	0.7	3195	2	US-08-335-844A-15	Sequence 15, Appl
C 315	15	0.7	2029	6	5457090-1	Sequence 3, Appl	C 389	15	0.7	3195	2	US-08-335-844A-15	Sequence 15, Appl
C 316	15	0.7	2029	6	5457090-1	Sequence 3, Appl	C 390	15	0.7	3195	2	US-08-335-844A-15	Sequence 15, Appl
C 317	15	0.7	2029	6	5457090-1	Sequence 3, Appl	C 391	15	0.7	3195	2	US-08-335-844A-15	Sequence 15, Appl
C 318	15	0.7	2029	6	5457090-1	Sequence 3, Appl	C 392	15	0.7	3195	2	US-08-335-844A-15	Sequence 15, Appl
C 319	15	0.7	2029	6	5457090-1	Sequence 3, Appl							

C 393	15	0.7	3876	5	PCT-US96-10782-1	Sequence 1, Appli	466	15	0.7	10825	3	US-08-652-265-7	Sequence 7, Appli
C 394	15	0.7	4046	1	US-08-706-936-1	Sequence 1, Appli	467	15	0.7	10825	3	US-08-834-497A-1	Sequence 1, Appli
C 395	15	0.7	4066	4	US-09-298-367B-3	Sequence 3, Appli	468	15	0.7	10825	3	US-08-834-497A-3	Sequence 3, Appli
C 396	15	0.7	4161	3	US-08-790-517-19	Sequence 1, Appli	469	15	0.7	10825	3	US-08-834-497A-5	Sequence 5, Appli
C 397	15	0.7	4161	3	US-08-790-517-19	Sequence 1, Appli	470	15	0.7	10825	3	US-08-834-497A-7	Sequence 7, Appli
C 398	15	0.7	4259	2	US-08-816-155B-2	Sequence 2, Appli	471	15	0.7	10825	4	US-09-503-444A-1	Sequence 1, Appli
C 399	15	0.7	4259	2	US-09-079-587-2	Sequence 2, Appli	472	15	0.7	10825	4	US-09-503-444A-3	Sequence 3, Appli
C 400	15	0.7	4315	6	US-08-882-046-3	Sequence 6, Appli	473	15	0.7	10825	4	US-09-503-444A-5	Sequence 5, Appli
C 401	15	0.7	4383	6	5175095-4	Patent No. 5175095	474	15	0.7	10825	4	US-09-503-444A-7	Sequence 7, Appli
C 402	15	0.7	4383	6	5177307-1	Patent No. 5177307	475	15	0.7	1103	4	US-09-101-886B-3	Sequence 7, Appli
C 403	15	0.7	4389	2	US-08-951-648-3	Sequence 3, Appli	476	15	0.7	1117	3	US-08-801-263A-4	Sequence 4, Appli
C 404	15	0.7	4389	2	US-09-174-437-3	Sequence 3, Appli	477	15	0.7	1117	3	US-09-102-248-4	Sequence 4, Appli
C 405	15	0.7	4464	2	US-08-400-159-7	Sequence 7, Appli	478	15	0.7	12225	2	US-08-822-445-11	Sequence 11, Appli
C 406	15	0.7	4483	3	US-08-611-729A-7	Sequence 7, Appli	479	15	0.7	12616	2	US-08-822-445-9	Sequence 9, Appli
C 407	15	0.7	4707	3	US-09-181-706-1	Sequence 1, Appli	480	15	0.7	13674	2	US-08-852-807-1	Sequence 1, Appli
C 408	15	0.7	4707	3	US-09-458-791-1	Sequence 1, Appli	481	15	0.7	14555	1	US-08-320-559-1	Sequence 1, Appli
C 409	15	0.7	4707	4	US-09-459-066-1	Sequence 1, Appli	482	15	0.7	14555	1	US-08-327-392-1	Sequence 1, Appli
C 410	15	0.7	4766	5	PCT-US93-07261-10	Sequence 10, Appli	483	15	0.7	14555	1	US-08-306-691B-55	Sequence 1, Appli
C 411	15	0.7	4808	1	US-08-351-413-17	Sequence 17, Appli	484	15	0.7	14555	1	US-08-545-860D-1	Sequence 55, Appli
C 412	15	0.7	4808	2	US-09-025-583-17	Sequence 17, Appli	485	15	0.7	14555	5	PCT-US94-04496-1	Sequence 1, Appli
C 413	15	0.7	4846	3	US-08-817-188-1	Sequence 1, Appli	486	15	0.7	14602	1	US-08-597-236-1	Sequence 1, Appli
C 414	15	0.7	5153	2	US-08-916-917-13	Sequence 8, Appli	487	15	0.7	14602	1	US-08-746-682A-1	Sequence 1, Appli
C 415	15	0.7	5162	2	US-09-225-170-13	Sequence 13, Appli	488	15	0.7	14796	4	US-08-975-080-35	Sequence 35, Appli
C 416	15	0.7	5162	3	US-09-225-170-13	Sequence 13, Appli	489	15	0.7	17410	1	US-08-147-023-3	Sequence 3, Appli
C 417	15	0.7	5162	4	US-09-298-367B-1	Sequence 1, Appli	490	15	0.7	17410	1	US-08-447-570-3	Sequence 3, Appli
C 418	15	0.7	5162	4	US-09-298-367B-4	Sequence 4, Appli	491	15	0.7	17410	1	US-08-449-700-3	Sequence 3, Appli
C 419	15	0.7	5227	2	US-08-996-306-3	Sequence 3, Appli	492	15	0.7	17410	2	US-08-449-699A-3	Sequence 3, Appli
C 420	15	0.7	5362	4	US-08-591-079-9	Sequence 4, Appli	493	15	0.7	17410	2	US-08-449-699A-3	Sequence 3, Appli
C 421	15	0.7	5393	2	US-08-342-930-1	Sequence 5, Appli	494	15	0.7	17415	5	PCT-US95-07349-1	Sequence 1, Appli
C 422	15	0.7	5455	1	US-09-068-195-21	Sequence 9, Appli	495	15	0.7	17415	5	PCT-US95-07349-1	Sequence 1, Appli
C 423	15	0.7	5560	3	US-08-882-046-1	Sequence 1, Appli	496	15	0.7	20137	4	US-09-262-773-206	Sequence 206, App
C 424	15	0.7	5560	3	US-08-882-046-1	Sequence 1, Appli	497	15	0.7	20138	4	US-09-262-773-9	Sequence 9, Appli
C 425	15	0.7	5560	1	US-08-104-072B-9	Sequence 9, Appli	498	15	0.7	23071	4	US-08-125-468-1	Sequence 20, App
C 426	15	0.7	5560	2	US-09-025-583-3	Sequence 3, Appli	499	15	0.7	30001	2	US-08-474-933-1	Sequence 1, Appli
C 427	15	0.7	5620	4	US-08-844-274-10	Sequence 10, Appli	500	15	0.7	30001	2	US-08-814-095-7	Sequence 7, Appli
C 428	15	0.7	5679	2	US-08-483-101-1	Sequence 1, Appli	501	15	0.7	35100	3	US-08-770-379-17	Sequence 17, Appli
C 429	15	0.7	5798	2	US-08-899-595-4	Sequence 4, Appli	502	15	0.7	35100	3	US-08-757-669A-17	Sequence 17, Appli
C 430	15	0.7	5822	3	US-08-899-595-5	Sequence 5, Appli	503	15	0.7	37895	1	US-08-375-709-1	Sequence 1, Appli
C 431	15	0.7	5822	3	US-07-867-106-2	Sequence 2, Appli	504	15	0.7	37895	1	US-08-752-929-1	Sequence 1, Appli
C 432	15	0.7	5864	3	US-08-894-440-4	Sequence 4, Appli	505	15	0.7	37895	3	US-09-090-793-1	Sequence 1, Appli
C 433	15	0.7	6291	3	US-08-817-188-4	Sequence 4, Appli	506	15	0.7	38956	3	US-09-320-878B-19	Sequence 19, Appli
C 434	15	0.7	6376	1	US-08-104-073-5	Sequence 5, Appli	507	15	0.7	48974	4	US-08-920-422-17	Sequence 17, Appli
C 435	15	0.7	6548	3	US-08-894-440-1	Sequence 1, Appli	508	15	0.7	48974	4	US-08-920-422-17	Sequence 17, Appli
C 436	15	0.7	6548	3	US-08-817-188-2	Sequence 2, Appli	509	15	0.7	48974	4	US-08-920-422-17	Sequence 17, Appli
C 437	15	0.7	6548	3	US-08-817-188-2	Sequence 2, Appli	510	15	0.7	48974	4	US-08-920-422-17	Sequence 17, Appli
C 438	15	0.7	6723	4	US-08-844-274-13	Sequence 13, Appli	511	15	0.7	48974	4	US-08-920-422-17	Sequence 17, Appli
C 439	15	0.7	6723	4	US-08-844-274-13	Sequence 13, Appli	512	15	0.7	48974	4	US-08-920-422-17	Sequence 17, Appli
C 440	15	0.7	6850	4	US-09-298-367B-2	Sequence 2, Appli	513	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli
C 441	15	0.7	7244	4	US-08-844-274-20	Sequence 20, Appli	514	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli
C 442	15	0.7	7560	4	US-08-844-274-20	Sequence 20, Appli	515	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli
C 443	15	0.7	7560	4	US-08-844-274-20	Sequence 20, Appli	516	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli
C 444	15	0.7	7742	2	US-08-619-554-1	Sequence 1, Appli	517	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli
C 445	15	0.7	7742	2	US-08-619-554-1	Sequence 1, Appli	518	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli
C 446	15	0.7	7833	1	US-08-882-704A-4	Sequence 4, Appli	519	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli
C 447	15	0.7	7833	1	US-08-882-704A-4	Sequence 4, Appli	520	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli
C 448	15	0.7	8438	4	US-07-945-283-1	Sequence 9, Appli	521	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli
C 449	15	0.7	8438	4	US-07-945-283-1	Sequence 9, Appli	522	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli
C 450	15	0.7	8874	3	US-08-894-344C-1	Sequence 1, Appli	523	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli
C 451	15	0.7	8874	3	US-08-894-344C-1	Sequence 1, Appli	524	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli
C 452	15	0.7	9421	2	US-08-370-319C-2	Sequence 2, Appli	525	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli
C 453	15	0.7	9421	2	US-08-370-319C-2	Sequence 2, Appli	526	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli
C 454	15	0.7	9421	4	US-09-224-834-2	Sequence 2, Appli	527	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli
C 455	15	0.7	9511	1	US-07-925-695-6	Sequence 6, Appli	528	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli
C 456	15	0.7	9511	1	US-07-925-695-6	Sequence 6, Appli	529	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli
C 457	15	0.7	9621	4	US-09-125-028-1	Sequence 1, Appli	530	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli
C 458	15	0.7	10163	2	US-08-659-251-1	Sequence 1, Appli	531	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli
C 459	15	0.7	10163	2	US-08-659-251-1	Sequence 1, Appli	532	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli
C 460	15	0.7	10163	5	PCT-US96-11445-1	Sequence 1, Appli	533	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli
C 461	15	0.7	10281	3	US-08-816-155B-1	Sequence 1, Appli	534	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli
C 462	15	0.7	10281	3	US-08-816-155B-1	Sequence 1, Appli	535	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli
C 463	15	0.7	10627	3	US-09-079-587-1	Sequence 12, Appli	536	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli
C 464	15	0.7	10825	3	US-08-652-265-1	Sequence 1, Appli	537	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli
C 465	15	0.7	10825	3	US-08-652-265-3	Sequence 3, Appli	538	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli
C 465	15	0.7	10825	3	US-08-652-265-5	Sequence 5, Appli	538	15	0.7	80161	3	US-08-781-891A-79	Sequence 79, Appli

539	14	0.7	28	1	US-08-171-718-55	Sequence 55, Appl	612	14	0.7	59	1	US-08-240-121-5	Sequence 5, Appl
540	14	0.7	29	3	US-08-478-087-55	Sequence 15, Appl	613	14	0.7	59	1	US-08-451-241-5	Sequence 5, Appl
541	14	0.7	28	3	US-08-617-484-15	Sequence 15, Appl	614	14	0.7	59	1	PCT-US92-10625-5	Sequence 5, Appl
542	14	0.7	30	1	US-08-113-646A-16	Sequence 16, Appl	615	14	0.7	59	5	PCT-US93-11297-5	Sequence 21, Appl
543	14	0.7	30	1	US-08-113-646A-34	Sequence 34, Appl	616	14	0.7	59	5	PCT-US93-11298-21	Sequence 6, Appl
544	14	0.7	30	3	US-09-192-048-11	Sequence 11, Appl	617	14	0.7	60	1	US-07-808-451-6	Sequence 6, Appl
545	14	0.7	30	3	US-09-192-048-12	Sequence 12, Appl	618	14	0.7	60	5	PCT-US92-10625-6	Sequence 2, Appl
546	14	0.7	32	1	US-08-113-646A-5	Sequence 5, Appl	619	14	0.7	61	1	US-08-207-547A-2	Sequence 2, Appl
547	14	0.7	32	1	US-08-628-665-13	Sequence 13, Appl	620	14	0.7	61	1	US-08-215-082-2	Sequence 2, Appl
548	14	0.7	33	1	US-08-113-646A-35	Sequence 35, Appl	621	14	0.7	61	2	US-08-702-652-2	Sequence 1, Appl
549	14	0.7	33	1	US-08-207-547A-7	Sequence 7, Appl	622	14	0.7	62	4	US-09-296-607-1	Sequence 2, Appl
550	14	0.7	33	1	US-08-215-082-7	Sequence 7, Appl	623	14	0.7	64	1	US-07-808-451-2	Sequence 1, Appl
551	14	0.7	33	2	US-08-702-652-7	Sequence 7, Appl	624	14	0.7	64	1	US-08-443-568B-33	Sequence 33, Appl
552	14	0.7	34	1	US-08-113-646A-36	Sequence 36, Appl	625	14	0.7	64	2	US-08-418-848A-11	Sequence 11, Appl
553	14	0.7	34	1	US-08-628-665-15	Sequence 15, Appl	626	14	0.7	64	5	PCT-US92-10625-2	Sequence 2, Appl
554	14	0.7	35	1	US-08-113-646A-37	Sequence 37, Appl	627	14	0.7	64	5	PCT-US94-06997-33	Sequence 33, Appl
555	14	0.7	35	4	US-09-232-479-2	Sequence 2, Appl	628	14	0.7	75	3	US-09-248-643-12	Sequence 12, Appl
556	14	0.7	36	1	US-08-113-646A-43	Sequence 43, Appl	629	14	0.7	79	1	US-08-091-569-20	Sequence 20, Appl
557	14	0.7	37	1	US-08-113-646A-38	Sequence 38, Appl	630	14	0.7	79	1	US-08-203-676-20	Sequence 20, Appl
558	14	0.7	37	1	US-08-338-992B-25	Sequence 25, Appl	631	14	0.7	79	2	US-08-822-238-20	Sequence 1, Appl
559	14	0.7	41	1	US-08-113-646A-39	Sequence 39, Appl	632	14	0.7	88	1	US-08-207-547A-1	Sequence 1, Appl
560	14	0.7	41	3	US-09-182-859-19	Sequence 19, Appl	633	14	0.7	88	1	US-08-215-082-1	Sequence 1, Appl
561	14	0.7	43	2	US-08-366-490-3	Sequence 3, Appl	634	14	0.7	88	2	US-08-702-652-1	Sequence 1, Appl
562	14	0.7	43	2	US-08-853-217-9	Sequence 9, Appl	635	14	0.7	91	1	US-08-120-827-101	Sequence 101, App
563	14	0.7	44	1	US-08-113-646A-40	Sequence 40, Appl	636	14	0.7	91	1	US-08-478-675-101	Sequence 101, App
564	14	0.7	44	3	US-08-860-483A-3	Sequence 3, Appl	637	14	0.7	91	1	US-08-210-222-12	Sequence 12, App
565	14	0.7	46	5	PCT-US93-02457-5	Sequence 5, Appl	638	14	0.7	100	1	US-08-091-569-19	Sequence 19, Appl
566	14	0.7	46	5	PCT-US93-11458-6	Sequence 6, Appl	639	14	0.7	100	2	US-08-203-676-19	Sequence 19, Appl
567	14	0.7	47	2	US-08-418-848A-62	Sequence 62, Appl	640	14	0.7	100	2	US-08-822-238-19	Sequence 4, Appl
568	14	0.7	48	2	US-08-477-553A-24	Sequence 34, Appl	641	14	0.7	129	3	US-08-937-610-4	Sequence 20, Appl
569	14	0.7	50	1	US-07-808-451-7	Sequence 7, Appl	642	14	0.7	177	2	US-09-032-684-20	Sequence 3, Appl
570	14	0.7	50	1	US-07-989-845-23	Sequence 23, Appl	643	14	0.7	178	1	US-08-471-119A-3	Sequence 6, Appl
571	14	0.7	50	1	US-07-989-844-7	Sequence 7, Appl	644	14	0.7	188	4	US-08-945-625A-6	Sequence 3, Appl
572	14	0.7	50	1	US-08-161-044-7	Sequence 7, Appl	645	14	0.7	219	2	US-08-858-830-3	Sequence 3, Appl
573	14	0.7	50	1	US-08-240-121-7	Sequence 7, Appl	646	14	0.7	219	2	US-08-858-834-3	Sequence 2, Appl
574	14	0.7	50	1	US-08-451-241-7	Sequence 7, Appl	647	14	0.7	225	1	US-07-789-917A-2	Sequence 2, Appl
575	14	0.7	50	1	US-08-443-568B-10	Sequence 30, Appl	648	14	0.7	225	1	US-08-444-142-3	Sequence 3, Appl
576	14	0.7	50	5	PCT-US92-10625-7	Sequence 7, Appl	649	14	0.7	233	1	US-08-444-131-3	Sequence 3, Appl
577	14	0.7	50	5	PCT-US93-11297-7	Sequence 7, Appl	650	14	0.7	233	1	US-07-764-655D-9	Sequence 9, Appl
578	14	0.7	50	5	PCT-US93-11298-23	Sequence 23, Appl	651	14	0.7	237	1	US-07-764-655D-9	Sequence 9, Appl
579	14	0.7	50	5	PCT-US94-06997-30	Sequence 30, Appl	652	14	0.7	237	1	US-08-208-486-6	Sequence 6, Appl
580	14	0.7	51	1	US-07-808-451-8	Sequence 8, Appl	653	14	0.7	239	1	US-08-171-718-102	Sequence 102, App
581	14	0.7	51	1	US-07-989-844-16	Sequence 16, Appl	654	14	0.7	243	3	US-08-478-087-102	Sequence 102, App
582	14	0.7	51	1	US-07-764-655D-1	Sequence 1, Appl	655	14	0.7	243	3	US-08-981-012A-1	Sequence 1, Appl
583	14	0.7	51	1	US-07-764-655D-10	Sequence 10, Appl	656	14	0.7	249	2	US-09-112-096-6	Sequence 6, Appl
584	14	0.7	51	1	US-08-110-663-4	Sequence 4, Appl	657	14	0.7	249	4	US-08-981-012A-1	Sequence 6, Appl
585	14	0.7	51	1	US-08-169-688-4	Sequence 4, Appl	658	14	0.7	252	2	US-08-981-012A-6	Sequence 13, Appl
586	14	0.7	51	1	US-08-240-121-16	Sequence 16, Appl	659	14	0.7	264	2	US-08-418-848A-13	Sequence 4, Appl
587	14	0.7	51	1	US-08-451-241-16	Sequence 16, Appl	660	14	0.7	266	2	US-09-019-201A-4	Sequence 197, App
588	14	0.7	51	1	US-08-110-664-4	Sequence 4, Appl	661	14	0.7	289	4	US-08-991-789A-197	Sequence 167, App
589	14	0.7	51	1	US-08-446-882-4	Sequence 4, Appl	662	14	0.7	294	2	US-08-853-217-28	Sequence 28, Appl
590	14	0.7	51	1	US-08-385-187A-4	Sequence 4, Appl	663	14	0.7	295	3	US-08-765-332-167	Sequence 1, Appl
591	14	0.7	51	1	US-08-470-108-4	Sequence 8, Appl	664	14	0.7	303	3	US-08-577-464-1	Sequence 26, Appl
592	14	0.7	51	1	US-08-443-568B-31	Sequence 31, Appl	665	14	0.7	324	2	US-08-853-217-26	Sequence 2, Appl
593	14	0.7	51	5	PCT-US92-10625-8	Sequence 8, Appl	666	14	0.7	329	2	US-08-849-536A-2	Sequence 29, Appl
594	14	0.7	51	5	PCT-US94-06997-31	Sequence 31, Appl	667	14	0.7	330	2	US-08-853-217-29	Sequence 5, Appl
595	14	0.7	51	5	PCT-US94-06997-32	Sequence 32, Appl	668	14	0.7	332	1	US-08-100-692-5	Sequence 5, Appl
596	14	0.7	55	1	US-07-808-451-1	Sequence 1, Appl	669	14	0.7	332	2	US-08-674-030-5	Sequence 8, Appl
597	14	0.7	55	1	US-07-989-845-17	Sequence 17, Appl	670	14	0.7	332	2	US-09-188-930-83	Sequence 55, Appl
598	14	0.7	55	1	US-07-989-844-1	Sequence 1, Appl	671	14	0.7	335	2	US-08-477-553A-55	Sequence 94, Appl
599	14	0.7	55	1	US-08-161-044-1	Sequence 1, Appl	672	14	0.7	342	2	US-08-691-814B-94	Sequence 10, Appl
600	14	0.7	55	1	US-08-240-121-1	Sequence 1, Appl	673	14	0.7	358	3	US-09-026-343-10	Sequence 34, Appl
601	14	0.7	55	1	US-08-113-646A-41	Sequence 41, Appl	674	14	0.7	365	1	US-09-026-343-24	Sequence 34, Appl
602	14	0.7	55	1	US-08-451-241-1	Sequence 1, Appl	675	14	0.7	365	1	US-08-222-177A-34	Sequence 1, Appl
603	14	0.7	55	1	US-08-443-568B-32	Sequence 32, Appl	676	14	0.7	393	1	US-07-790-199-1	Sequence 34, Appl
604	14	0.7	55	5	PCT-US92-10625-1	Sequence 1, Appl	677	14	0.7	396	1	US-08-470-179-34	Sequence 17, Appl
605	14	0.7	55	5	PCT-US93-11297-1	Sequence 1, Appl	678	14	0.7	400	4	US-08-825-852-17	Sequence 17, Appl
606	14	0.7	55	5	PCT-US93-11298-17	Sequence 17, Appl	679	14	0.7	400	4	US-09-052-888-17	Sequence 8, Appl
607	14	0.7	55	5	PCT-US94-06997-32	Sequence 32, Appl	680	14	0.7	408	4	US-09-293-395-8	Sequence 20, Appl
608	14	0.7	59	1	US-07-808-451-5	Sequence 5, Appl	681	14	0.7	414	2	US-08-766-439-20	Sequence 20, Appl
609	14	0.7	59	1	US-07-989-845-21	Sequence 21, Appl	682	14	0.7	414	2	US-08-766-439-21	Sequence 21, Appl
610	14	0.7	59	1	US-07-989-844-5	Sequence 5, Appl	683	14	0.7	414	2	US-08-766-439-22	Sequence 22, Appl
611	14	0.7	59	1	US-08-161-044-5	Sequence 5, Appl	684	14	0.7	414	2	US-08-766-439-23	Sequence 23, Appl

C 685	14	0.7	417	2	US-08-245-511-11	Sequence 11, Appl	758	14	0.7	601	2	US-08-184-009-168	Sequence 168, App
C 686	14	0.7	420	2	US-08-600-993A-11	Sequence 11, Appl	759	14	0.7	601	2	US-08-458-356-168	Sequence 168, App
C 687	14	0.7	420	2	US-08-189-256A-20	Sequence 20, Appl	760	14	0.7	606	2	US-08-676-279-51	Sequence 51, Appl
C 688	14	0.7	421	3	US-08-961-083-99	Sequence 99, Appl	761	14	0.7	612	3	US-09-280-799-208	Sequence 208, App
C 689	14	0.7	422	1	US-08-470-179-73	Sequence 73, Appl	762	14	0.7	616	4	US-09-276-531-39	Sequence 39, Appl
C 690	14	0.7	426	3	US-08-779-764A-7	Sequence 6, Appl	763	14	0.7	629	1	US-08-592-126-135	Sequence 135, App
C 691	14	0.7	426	3	US-08-779-764A-7	Sequence 7, Appl	764	14	0.7	631	4	US-09-130-242-3	Sequence 3, Appl
C 692	14	0.7	426	3	US-08-779-764A-8	Sequence 8, Appl	765	14	0.7	634	1	US-07-865-662F-3	Sequence 3, Appl
C 693	14	0.7	426	3	US-08-779-764A-9	Sequence 9, Appl	766	14	0.7	634	4	US-08-374-219B-3	Sequence 3, Appl
C 694	14	0.7	428	3	US-08-369-822C-5	Sequence 5, Appl	767	14	0.7	649	4	US-08-998-416-184	Sequence 184, App
C 695	14	0.7	428	3	US-08-582-776C-5	Sequence 5, Appl	768	14	0.7	655	4	US-08-998-416-848	Sequence 848, App
C 696	14	0.7	428	3	US-08-434-831B-5	Sequence 5, Appl	769	14	0.7	663	1	US-07-624-313-3	Sequence 718, App
C 697	14	0.7	432	2	US-08-967-101-70	Sequence 70, Appl	770	14	0.7	668	6	5498529-5	Sequence 3, Appl
C 698	14	0.7	435	2	US-08-352-541-70	Sequence 70, Appl	771	14	0.7	672	1	US-07-790-199-3	Sequence 3, Appl
C 699	14	0.7	435	3	US-09-124-698-70	Sequence 70, Appl	772	14	0.7	680	4	US-08-896-164-39	Sequence 39, Appl
C 700	14	0.7	435	4	US-09-127-480-70	Sequence 70, Appl	773	14	0.7	682	4	US-08-896-164-39	Sequence 136, App
C 701	14	0.7	441	4	US-08-496-841C-70	Sequence 70, Appl	774	14	0.7	696	1	US-08-592-126-136	Sequence 36, Appl
C 702	14	0.7	457	1	US-08-482-271-2	Sequence 2, Appl	775	14	0.7	702	1	US-08-284-784-36	Sequence 870, App
C 703	14	0.7	457	4	US-09-158-710-17	Sequence 17, Appl	776	14	0.7	702	2	US-08-854-811-36	Sequence 785, App
C 704	14	0.7	457	4	US-08-924-755-7	Sequence 17, Appl	777	14	0.7	707	4	US-08-998-416-870	Sequence 787, App
C 705	14	0.7	458	2	US-08-248-335-7	Sequence 7, Appl	778	14	0.7	710	4	US-08-998-416-785	Sequence 48, App
C 706	14	0.7	458	3	PCT-US93-01676A-7	Sequence 7, Appl	779	14	0.7	711	1	US-08-307-499-49	Sequence 7, Appl
C 707	14	0.7	459	5	PCT-US93-01676A-8	Sequence 8, Appl	780	14	0.7	711	3	US-08-493-071-7	Sequence 49, Appl
C 708	14	0.7	485	1	US-07-989-845-29	Sequence 29, Appl	781	14	0.7	711	6	US-09-299-268-49	Sequence 49, Appl
C 709	14	0.7	485	1	US-07-989-844-13	Sequence 13, Appl	782	14	0.7	711	6	5169941-24	Sequence 16, App
C 710	14	0.7	485	1	US-08-110-663-1	Sequence 1, Appl	783	14	0.7	713	4	US-08-998-416-616	Sequence 616, App
C 711	14	0.7	485	1	US-08-169-688-1	Sequence 1, Appl	784	14	0.7	714	3	US-08-470-535-13	Sequence 832, App
C 712	14	0.7	485	1	US-08-240-121-13	Sequence 13, Appl	785	14	0.7	718	4	US-08-998-416-832	Sequence 832, App
C 713	14	0.7	485	1	US-08-451-241-13	Sequence 13, Appl	786	14	0.7	723	6	5223425-12	Sequence 6, Appl
C 714	14	0.7	485	1	US-08-110-664-1	Sequence 1, Appl	787	14	0.7	727	1	US-07-885-970A-6	Sequence 6, Appl
C 715	14	0.7	485	1	US-08-446-882-1	Sequence 1, Appl	788	14	0.7	727	1	US-08-298-687A-6	Sequence 5, Appl
C 716	14	0.7	485	1	US-08-385-187A-1	Sequence 1, Appl	789	14	0.7	727	1	US-08-998-416-966	Sequence 5, Appl
C 717	14	0.7	485	1	US-08-470-108-1	Sequence 1, Appl	790	14	0.7	727	1	US-08-998-416-966	Sequence 6, Appl
C 718	14	0.7	485	5	PCT-US93-11297-13	Sequence 13, Appl	791	14	0.7	727	2	US-08-787-335-5	Sequence 6, Appl
C 719	14	0.7	485	5	PCT-US93-11298-29	Sequence 29, Appl	792	14	0.7	727	2	US-08-998-416-1031	Sequence 1031, App
C 720	14	0.7	485	2	US-08-967-101-92	Sequence 92, Appl	793	14	0.7	729	4	US-08-451-405A-2	Sequence 2, Appl
C 721	14	0.7	495	2	US-08-592-541-92	Sequence 92, Appl	794	14	0.7	731	2	US-08-316-650-15	Sequence 15, Appl
C 722	14	0.7	495	2	US-09-124-698-92	Sequence 92, Appl	795	14	0.7	731	2	US-08-998-416-966	Sequence 15, Appl
C 723	14	0.7	495	4	US-09-127-480-92	Sequence 92, Appl	796	14	0.7	731	5	PCT-US95-02251-15	Sequence 15, Appl
C 724	14	0.7	495	4	US-08-966-269-3	Sequence 3, Appl	797	14	0.7	732	4	US-08-998-416-938	Sequence 3, Appl
C 725	14	0.7	495	4	US-07-918-953-7	Sequence 7, Appl	798	14	0.7	732	4	US-08-963-743-2	Sequence 2, Appl
C 726	14	0.7	495	4	US-08-081-661-7	Sequence 11, Appl	799	14	0.7	732	4	US-08-998-416-1035	Sequence 1035, App
C 727	14	0.7	497	1	US-08-489-066A-1	Sequence 1, Appl	800	14	0.7	737	1	US-07-808-451-11	Sequence 11, Appl
C 728	14	0.7	498	2	US-08-631-328-54	Sequence 11, Appl	801	14	0.7	749	1	US-07-989-845-27	Sequence 27, Appl
C 729	14	0.7	504	3	US-08-965-903B-14	Sequence 14, Appl	802	14	0.7	757	1	US-07-989-845-27	Sequence 11, Appl
C 730	14	0.7	506	3	US-08-966-269-3	Sequence 3, Appl	803	14	0.7	757	1	US-08-161-044-11	Sequence 11, Appl
C 731	14	0.7	510	1	US-07-918-953-7	Sequence 7, Appl	804	14	0.7	757	1	US-08-240-121-11	Sequence 11, Appl
C 732	14	0.7	510	1	US-08-081-661-7	Sequence 11, Appl	805	14	0.7	757	1	US-08-451-241-11	Sequence 11, Appl
C 733	14	0.7	519	3	US-08-860-165-11	Sequence 11, Appl	806	14	0.7	757	1	PCT-US92-10625-11	Sequence 11, Appl
C 734	14	0.7	528	1	US-08-044-506B-31	Sequence 31, Appl	807	14	0.7	757	5	PCT-US92-10625-11	Sequence 11, Appl
C 735	14	0.7	533	6	5482709-5	Sequence 5, Appl	808	14	0.7	757	5	PCT-US93-11297-11	Sequence 11, Appl
C 736	14	0.7	540	1	US-07-990-965-1	Sequence 1, Appl	809	14	0.7	757	5	PCT-US93-11297-11	Sequence 11, Appl
C 737	14	0.7	542	4	US-09-305-639-5	Sequence 5, Appl	810	14	0.7	757	5	PCT-US93-11297-11	Sequence 11, Appl
C 738	14	0.7	543	6	5273901-6	Sequence 6, Appl	811	14	0.7	757	5	PCT-US93-11297-11	Sequence 11, Appl
C 739	14	0.7	552	2	US-08-721-488-2	Sequence 2, Appl	812	14	0.7	757	5	PCT-US93-11297-11	Sequence 11, Appl
C 740	14	0.7	552	2	US-08-731-722-7	Sequence 7, Appl	813	14	0.7	757	5	PCT-US93-11297-11	Sequence 11, Appl
C 741	14	0.7	567	3	US-08-995-960-8	Sequence 8, Appl	814	14	0.7	770	1	US-07-953-230A-1	Sequence 1, Appl
C 742	14	0.7	570	2	US-08-257-784A-12	Sequence 12, Appl	815	14	0.7	776	1	US-08-216-233C-1	Sequence 1, Appl
C 743	14	0.7	570	2	US-08-489-066A-1	Sequence 1, Appl	816	14	0.7	776	3	US-09-090-947-19	Sequence 19, Appl
C 744	14	0.7	570	3	US-08-907-706-4	Sequence 4, Appl	817	14	0.7	782	4	US-08-998-416-403	Sequence 403, App
C 745	14	0.7	570	3	US-08-489-072A-1	Sequence 1, Appl	818	14	0.7	786	1	US-08-644-664B-25	Sequence 29, Appl
C 746	14	0.7	570	5	PCT-US94-05085A-1	Sequence 1, Appl	819	14	0.7	786	2	US-08-761-277A-29	Sequence 29, Appl
C 747	14	0.7	570	5	PCT-US94-05085A-1	Sequence 1, Appl	820	14	0.7	787	3	US-08-961-083-97	Sequence 97, Appl
C 748	14	0.7	586	2	US-08-849-536A-1	Sequence 1, Appl	821	14	0.7	790	1	US-08-117-083-7	Sequence 7, Appl
C 749	14	0.7	590	1	US-08-466-033-116	Sequence 116, App	822	14	0.7	790	1	US-08-998-416-459	Sequence 459, App
C 750	14	0.7	590	1	US-08-444-733-116	Sequence 116, App	823	14	0.7	794	4	US-08-493-071-10	Sequence 10, Appl
C 751	14	0.7	590	2	US-08-464-134-116	Sequence 116, App	824	14	0.7	795	3	US-08-929-302-3	Sequence 3, Appl
C 752	14	0.7	590	2	US-08-461-361-116	Sequence 116, App	825	14	0.7	800	2	US-09-038-014-3	Sequence 3, Appl
C 753	14	0.7	592	4	US-08-485-910-116	Sequence 80, Appl	826	14	0.7	800	2	US-08-417-210A-93	Sequence 93, Appl
C 754	14	0.7	592	4	US-08-896-164-80	Sequence 80, Appl	827	14	0.7	801	3	US-08-860-165-9	Sequence 9, Appl
C 755	14	0.7	594	1	US-08-404-732A-3	Sequence 163, App	828	14	0.7				
C 756	14	0.7	598	1	US-08-184-009-163	Sequence 163, App	829	14	0.7				
C 757	14	0.7	598	2	US-08-458-356-163	Sequence 163, App	830	14	0.7				



831	14	0.7	810	2	US-08-611-880-3	Sequence 3, Appl	c 904	14	0.7	1041	1	US-08-747-240-1	Sequence 1, Appl
c 832	14	0.7	811	3	US-08-961-083-205	Sequence 205, App	905	14	0.7	1044	4	US-09-094-557-4	Sequence 4, Appl
c 833	14	0.7	813	1	US-08-182-114-2	Sequence 2, Appl	c 906	14	0.7	1050	4	US-09-137-853-1	Sequence 1, Appl
c 834	14	0.7	834	1	US-08-690-457-8	Sequence 8, Appl	907	14	0.7	1051	6	5498529-7	Patent No. 5498529
c 835	14	0.7	834	2	US-08-628-187-8	Sequence 8, Appl	908	14	0.7	1051	6	5498529-7	Sequence 7, Appl
c 836	14	0.7	840	1	US-08-595-559-4	Sequence 4, Appl	909	14	0.7	1052	2	US-08-403-852D-10	Sequence 10, Appl
c 837	14	0.7	846	1	US-07-953-230A-5	Sequence 5, Appl	910	14	0.7	1052	4	US-08-510-646B-10	Sequence 10, Appl
c 838	14	0.7	847	4	US-08-998-416-306	Sequence 306, App	911	14	0.7	1052	4	US-08-454-196-16	Sequence 16, Appl
c 839	14	0.7	850	1	US-08-247-809A-11	Sequence 11, Appl	c 912	14	0.7	1079	3	US-09-064-033-16	Sequence 16, Appl
c 840	14	0.7	850	2	US-08-711-728-11	Sequence 11, Appl	c 913	14	0.7	1089	3	US-09-195-666A-10	Sequence 10, Appl
c 841	14	0.7	856	4	US-09-318-661-3	Sequence 3, Appl	c 914	14	0.7	1101	3	US-08-746-883-3	Sequence 3, Appl
c 842	14	0.7	856	4	US-09-318-661-6	Sequence 6, Appl	c 915	14	0.7	1103	3	US-08-617-860B-5	Sequence 5, Appl
c 843	14	0.7	864	1	US-08-690-457-7	Sequence 7, Appl	916	14	0.7	1106	3	US-08-844-055-1	Sequence 1, Appl
c 844	14	0.7	864	2	US-08-628-187-7	Sequence 7, Appl	917	14	0.7	1107	3	US-09-006-849-1	Sequence 1, Appl
c 845	14	0.7	864	3	US-08-906-769-138	Sequence 18, App	c 918	14	0.7	1110	1	US-07-941-523-20	Sequence 20, Appl
c 846	14	0.7	864	3	US-08-906-616-138	Sequence 18, App	c 919	14	0.7	1110	1	US-08-202-186-15	Sequence 15, Appl
c 847	14	0.7	864	3	US-08-639-075A-138	Sequence 138, App	c 920	14	0.7	1111	1	US-08-709-209-87	Sequence 87, Appl
c 848	14	0.7	864	4	US-09-012-431-138	Sequence 138, App	c 921	14	0.7	1112	3	US-08-735-587-25	Sequence 25, Appl
c 849	14	0.7	864	4	US-09-012-692-138	Sequence 138, App	c 922	14	0.7	1113	3	US-08-891-322-5	Sequence 5, Appl
c 850	14	0.7	864	4	US-08-906-613-138	Sequence 138, App	c 923	14	0.7	1113	3	US-09-277-019-5	Sequence 63, Appl
c 851	14	0.7	864	1	US-08-690-457-6	Sequence 6, Appl	c 924	14	0.7	1116	4	US-08-683-262B-63	Sequence 2, Appl
c 852	14	0.7	867	2	US-08-628-187-6	Sequence 6, Appl	c 925	14	0.7	1116	4	US-08-683-262B-63	Sequence 2, Appl
c 853	14	0.7	868	3	US-08-889-502-20	Sequence 20, Appl	c 926	14	0.7	1120	3	US-08-832-180-6	Sequence 6, Appl
c 854	14	0.7	876	3	US-09-273-742-1	Sequence 1, Appl	c 927	14	0.7	1120	3	US-08-199-219-7	Sequence 7, Appl
c 855	14	0.7	879	1	US-08-182-114-1	Sequence 1, Appl	c 928	14	0.7	1132	3	US-08-900-565-2	Sequence 2, Appl
c 856	14	0.7	886	1	US-07-923-692C-3	Sequence 3, Appl	c 929	14	0.7	1132	3	US-08-105-483-87	Sequence 87, Appl
c 857	14	0.7	886	1	US-08-184-237-3	Sequence 3, Appl	c 930	14	0.7	1142	1	US-08-709-209-87	Sequence 87, Appl
c 858	14	0.7	886	2	US-08-482-920-3	Sequence 3, Appl	c 931	14	0.7	1142	1	US-08-458-101-87	Sequence 15, Appl
c 859	14	0.7	886	2	US-08-484-341-3	Sequence 3, Appl	c 932	14	0.7	1147	1	US-08-417-103-15	Sequence 1, Appl
c 860	14	0.7	895	2	US-08-924-759-23	Sequence 23, Appl	c 933	14	0.7	1150	3	US-09-046-086-1	Sequence 1, Appl
c 861	14	0.7	895	3	US-09-248-335-23	Sequence 23, Appl	c 934	14	0.7	1150	3	US-08-995-159-1	Sequence 1, Appl
c 862	14	0.7	900	4	US-08-879-096-4	Sequence 4, Appl	c 935	14	0.7	1160	6	5169941-2	Patent No. 5169941
c 863	14	0.7	900	4	US-08-872-961A-2	Sequence 2, Appl	c 936	14	0.7	1160	6	5169941-2	Sequence 11, Appl
c 864	14	0.7	904	3	US-09-231-258-2	Sequence 2, Appl	c 937	14	0.7	1164	1	US-08-445-640-11	Sequence 11, Appl
c 865	14	0.7	904	3	US-08-854-811-42	Sequence 42, Appl	c 938	14	0.7	1164	3	US-08-447-314-11	Sequence 11, Appl
c 866	14	0.7	915	1	US-08-854-811-42	Sequence 42, Appl	c 939	14	0.7	1164	3	US-08-445-640-11	Sequence 11, Appl
c 867	14	0.7	934	2	US-08-731-722-6	Sequence 6, Appl	c 940	14	0.7	1173	3	US-08-445-640-11	Sequence 9, Appl
c 868	14	0.7	934	2	US-08-448-600-2	Sequence 2, Appl	c 941	14	0.7	1173	3	US-08-403-917A-24	Sequence 24, Appl
c 869	14	0.7	940	1	US-08-067-800-7	Sequence 7, Appl	c 942	14	0.7	1173	3	US-08-403-917A-24	Sequence 24, Appl
c 870	14	0.7	942	1	US-08-021-537-2	Sequence 2, Appl	c 943	14	0.7	1173	3	US-08-403-917A-24	Sequence 24, Appl
c 871	14	0.7	959	4	US-09-067-800-7	Sequence 7, Appl	c 944	14	0.7	1173	3	US-08-403-917A-24	Sequence 24, Appl
c 872	14	0.7	966	2	US-08-514-014-7	Sequence 7, Appl	c 945	14	0.7	1183	2	US-08-731-722-8	Sequence 8, Appl
c 873	14	0.7	966	2	US-08-833-823-7	Sequence 7, Appl	c 946	14	0.7	1183	2	US-08-731-722-8	Sequence 8, Appl
c 874	14	0.7	966	2	US-08-932-978-1	Sequence 2, Appl	c 947	14	0.7	1186	2	US-08-731-722-5	Sequence 5, Appl
c 875	14	0.7	966	2	US-08-932-978-1	Sequence 2, Appl	c 948	14	0.7	1186	2	US-08-731-722-5	Sequence 5, Appl
c 876	14	0.7	966	2	US-08-932-978-1	Sequence 2, Appl	c 949	14	0.7	1186	2	US-08-731-722-5	Sequence 5, Appl
c 877	14	0.7	966	2	US-08-932-978-1	Sequence 2, Appl	c 950	14	0.7	1188	2	US-08-624-601-6	Sequence 6, Appl
c 878	14	0.7	999	6	5514567-5	Patent No. 5514567	c 951	14	0.7	1188	2	US-08-624-601-6	Sequence 6, Appl
c 879	14	0.7	1000	1	US-07-965-274-3	Sequence 3, Appl	c 952	14	0.7	1197	1	US-08-624-601-6	Sequence 6, Appl
c 880	14	0.7	1000	4	US-09-032-523-4	Sequence 4, Appl	c 953	14	0.7	1197	1	US-08-624-601-6	Sequence 6, Appl
c 881	14	0.7	1008	1	US-08-252-966B-15	Sequence 15, Appl	c 954	14	0.7	1197	1	US-08-624-601-6	Sequence 6, Appl
c 882	14	0.7	1010	4	US-09-011-526-3	Sequence 3, Appl	c 955	14	0.7	1204	1	US-08-628-291-11	Sequence 11, Appl
c 883	14	0.7	1020	2	US-07-757-606B-3	Sequence 7, Appl	c 956	14	0.7	1204	1	US-08-628-291-11	Sequence 11, Appl
c 884	14	0.7	1020	2	US-07-757-606B-3	Sequence 7, Appl	c 957	14	0.7	1215	6	5169941-7	Patent No. 5169941
c 885	14	0.7	1020	2	US-08-714-918-70	Sequence 70, Appl	c 958	14	0.7	1215	6	5169941-7	Sequence 4, Appl
c 886	14	0.7	1021	4	US-09-265-315-70	Sequence 70, Appl	c 959	14	0.7	1218	2	US-08-731-722-4	Sequence 4, Appl
c 887	14	0.7	1021	4	US-09-265-315-70	Sequence 70, Appl	c 960	14	0.7	1220	6	5204252-1	Patent No. 5204252
c 888	14	0.7	1021	4	US-09-266-417-70	Sequence 70, Appl	c 961	14	0.7	1227	2	US-08-080-386-1	Sequence 1, Appl
c 889	14	0.7	1023	2	US-08-307-499-53	Sequence 53, Appl	c 962	14	0.7	1227	2	US-08-390-000A-1	Sequence 1, Appl
c 890	14	0.7	1023	2	US-08-757-653-175	Sequence 175, App	c 963	14	0.7	1227	2	US-08-772-440-3	Sequence 3, Appl
c 891	14	0.7	1023	2	US-08-823-516-78	Sequence 78, App	c 964	14	0.7	1235	4	US-09-587-066-1	Sequence 1, Appl
c 892	14	0.7	1023	3	US-08-759-038-114	Sequence 114, App	c 965	14	0.7	1235	4	US-09-587-066-1	Sequence 1, Appl
c 893	14	0.7	1023	4	US-08-284-784-31	Sequence 31, App	c 966	14	0.7	1235	4	US-09-587-066-1	Sequence 1, Appl
c 894	14	0.7	1023	4	US-08-284-784-31	Sequence 31, App	c 967	14	0.7	1235	4	US-09-587-066-1	Sequence 1, Appl
c 895	14	0.7	1032	1	US-08-854-811-31	Sequence 31, App	c 968	14	0.7	1235	4	US-09-587-066-1	Sequence 1, Appl
c 896	14	0.7	1032	2	US-08-667-809B-5	Sequence 5, App	c 969	14	0.7	1235	4	US-09-587-066-1	Sequence 1, Appl
c 897	14	0.7	1032	2	US-08-667-809B-5	Sequence 5, App	c 970	14	0.7	1235	4	US-09-587-066-1	Sequence 1, Appl
c 898	14	0.7	1035	2	US-08-417-210A-111	Sequence 111, App	c 971	14	0.7	1259	1	US-08-844-064-5	Sequence 5, Appl
c 899	14	0.7	1040	2	US-08-247-809A-1	Sequence 1, Appl	c 972	14	0.7	1259	1	US-08-844-064-5	Sequence 5, Appl
c 900	14	0.7	1040	2	US-08-711-728-1	Sequence 1, Appl	c 973	14	0.7	1259	1	US-08-844-064-5	Sequence 5, Appl
c 901	14	0.7	1041	1	US-08-213-403-1	Sequence 1, Appl	c 974	14	0.7	1274	2	US-08-766-439-28	Sequence 28, Appl
c 902	14	0.7	1041	1	US-08-458-077-1	Sequence 1, Appl	c 975	14	0.7	1274	2	US-08-766-439-28	Sequence 28, Appl
c 903	14	0.7	1041	1	US-08-460-741-1	Sequence 1, Appl	c 976	14	0.7	1275	2	US-08-958-642-1	Sequence 1, Appl



```

; TITLE OF INVENTION: REGULATION OF GENE EXPRESSION BY
; EMPLOYING TRANSLATIONAL INHIBITION OF MRNA UTILIZING
; INTERFERING COMPLEMENTARY MRNA
; NUMBER OF SEQUENCES: 7
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/436,598
; FILING DATE: 15-NOV-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 300,741
; FILING DATE: 23-JAN-1989
; APPLICATION NUMBER: 228,852
; FILING DATE: 03-AUG-1988
; APPLICATION NUMBER: 543,528
; FILING DATE: 20-OCT-1983
; SEQ ID NO:2:
; LENGTH: 174
; 5190931-2

```

```

Query Match          0.9%; Score 19; DB 6; Length 174;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2007 aaataagtaatgatgat 2025
Db 103 AAATAAGTTAATGATGAT 85

```

```

RESULT 3
US-08-189-256A-20
; Sequence 20, Application US/08189256A
; Patent No. 5877402
; GENERAL INFORMATION:
; APPLICANT: Maliqa, Pal
; APPLICANT: Svab, Zora
; APPLICANT: Staub, Jeffrey
; APPLICANT: Zoubenko, Oleg V.
; APPLICANT: Allison, Lori A.
; APPLICANT: Carrier, Helaine
; APPLICANT: Kanevski, Ivan
; TITLE OF INVENTION: DNA Constructs and Methods for Stably
; TITLE OF INVENTION: Transforming Plastids of Multicellular Plants and
; TITLE OF INVENTION: Expressing Recombinant Proteins Therein
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/189,256A
; FILING DATE: 31-JAN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/111,398
; FILING DATE: 25-AUG-1993
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/518,763
; FILING DATE: 01-MAY-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 20:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 420 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; US-08-189-256A-20

```

```

Query Match          0.9%; Score 19; DB 2; Length 420;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1253 aaaaaaataatgatg 1271
Db 299 AAAAAAATCAATTTTG 317

```

```

RESULT 4
5190931-1
; Patent No. 5190931
; APPLICANT: INOUYE, MASAYORI
; TITLE OF INVENTION: REGULATION OF GENE EXPRESSION BY
; EMPLOYING TRANSLATIONAL INHIBITION OF MRNA UTILIZING
; INTERFERING COMPLEMENTARY MRNA
; NUMBER OF SEQUENCES: 7
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/436,598
; FILING DATE: 15-NOV-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 300,741
; FILING DATE: 23-JAN-1989
; APPLICATION NUMBER: 228,852
; FILING DATE: 03-AUG-1988
; APPLICATION NUMBER: 543,528
; FILING DATE: 20-OCT-1983
; SEQ ID NO:1:
; LENGTH: 500
; 5190931-1

```

```

Query Match          0.9%; Score 19; DB 6; Length 500;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2007 aaataagtaatgatgat 2025
Db 135 aaataagtaatgatgat 153

```

```

RESULT 5
US-08-189-256A-27
; Sequence 27, Application US/08189256A
; Patent No. 5877402
; GENERAL INFORMATION:
; APPLICANT: Maliqa, Pal
; APPLICANT: Svab, Zora
; APPLICANT: Staub, Jeffrey
; APPLICANT: Zoubenko, Oleg V.
; APPLICANT: Allison, Lori A.
; APPLICANT: Carrier, Helaine
; APPLICANT: Kanevski, Ivan
; TITLE OF INVENTION: DNA Constructs and Methods for Stably
; TITLE OF INVENTION: Transforming Plastids of Multicellular Plants and
; TITLE OF INVENTION: Expressing Recombinant Proteins Therein
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA

```

```

? COUNTRY: USA
? ZIP: 19103-2307
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/189,256A
? FILING DATE: 31-JAN-1994
? CLASSIFICATION: 435
? PRIORITY APPLICATION DATA:
? APPLICATION NUMBER: US 08/111,398
? FILING DATE: 25-AUG-1993
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/518,763
? FILING DATE: 01-MAY-1990
? ATTORNEY/AGENT INFORMATION:
? NAME: Reed, Janet E.
? REGISTRATION NUMBER: 36,252
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (215) 563-4100
? TELEFAX: (215) 563-4044
? INFORMATION FOR SEQ ID NO: 27:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1416 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? HYPOTHEICAL: NO
? ANTI-SENSE: NO
?
US-08-189-256A-27

Query Match      0.9%; Score 19; DB 2; Length 1416;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1253 aaaaaaaccaatttg 1271
Db 1295 AAAAAAATCAATTG 1313

RESULT
US-08-998-416-1022
? Sequence 1022, Application US/08998416
? Patent No. 6239264
? GENERAL INFORMATION:
? APPLICANT: Philippen, Peter
? APPLICANT: Pohlmann, Rainer
? APPLICANT: Steiner, Sabine
? APPLICANT: Mohr, Christine
? APPLICANT: Wendland, Jurgen
? APPLICANT: Knechtle, Philipp
? APPLICANT: Reibschung, Corinne
? TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPPII
? NUMBER OF SEQUENCES: 1152
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: No. 6239264artis Corporation
? STREET: 3054 Cornwallis Road
? CITY: Research Triangle Park
? STATE: No. 6239264th Carolina
? COUNTRY: USA
? ZIP: 27709
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/998,416
```

```

? FILING DATE: 24-DEC-1997
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: CH 0016/97
? FILING DATE: 31-DEC-1996
? ATTORNEY/AGENT INFORMATION:
? NAME: Meigs, J. Timothy
? REGISTRATION NUMBER: 36,241
? REFERENCE/DOCKET NUMBER: PP/5-30306/A/CGC1976
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 919-541-8587
? TELEFAX: 919-541-8689
? INFORMATION FOR SEQ ID NO: 1022:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 643 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? ORIGINAL SOURCE:
? ORGANISM: PAG1626RP
?
US-08-998-416-1022

Query Match      0.9%; Score 18; DB 4; Length 643;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 gcgcgcgcgcgcgcgcgc 97
Db 465 GCGCGCGCGCGCGCACCA 482

RESULT
US-09-129-888-1
? Sequence 1, Application US/09129888B
? Patent No. 6063912
? GENERAL INFORMATION:
? APPLICANT: CHUN, Jong Yoon
? APPLICANT: HAN, Yun Jeong
? TITLE OF INVENTION: Placenta trophoblast-specific gene
? FILE REFERENCE: 1942/29
? CURRENT APPLICATION NUMBER: US/09/129,888B
? CURRENT FILING DATE: 1998-08-06
? NUMBER OF SEQ ID NOS: 6
? SOFTWARE: WordPerfect 6.1 Windows
? SEQ ID NO 1
? LENGTH: 877
? TYPE: DNA
? ORGANISM: mouse
?
US-09-129-888-1

Query Match      0.9%; Score 18; DB 3; Length 877;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 861 tgggtggaatgcagtggt 878
Db 124 tgggtggaatgcagtggt 141

RESULT
US-08-718-738-1/c
? Sequence 1, Application US/08718738
? Patent No. 6013469
? GENERAL INFORMATION:
? APPLICANT: KUNSCH, CHARLES A
? APPLICANT: CHOPRA, ARYIND
? APPLICANT: ROSEN, CRAIG A
? TITLE OF INVENTION: HUMAN B-CELL TRANSLOCATION GENES-2 AND 3
? NUMBER OF SEQUENCES: 18
? CORRESPONDENCE ADDRESS:
```

```

ADDRESS: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/718,738
FILING DATE: 18-SEP-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0730003/ENS/KMT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1843 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 27..1061
US-08-718-738-1

Query Match          0.9%; Score 18; DB 3; Length 1843;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1248 ttataaaaaaatcaa 1265
Db 1338 TTTATATAAAAAATCAA 1321

RESULT 9
PCT-US95-03323A-1/C
Sequence 1, Application PC/TUS9503323A
GENERAL INFORMATION:
APPLICANT: KUNSCHE, ET AL.
TITLE OF INVENTION: Human B-Cell Translocation Genes-2 and 3
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: CECCHI, BYRNE, BAIN, GILFILLAN,
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/03323A
FILING DATE: Submitted herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.

```

```

REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-262
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1843 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
PCT-US95-03323A-1

Query Match          0.9%; Score 18; DB 5; Length 1843;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1248 ttataaaaaaatcaa 1265
Db 1338 TTTATATAAAAAATCAA 1321

RESULT 10
US-08-574-959A-1
Sequence 1, Application US/08574959A
Patent No. 5962224
GENERAL INFORMATION:
APPLICANT: Jaekyoon Shin, Insil Joung, Ratna K. Vadlamudi
APPLICANT: and Jack L. Strominger
TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES
TITLE OF INVENTION: AND USES THEREFOR
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, Suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/574,959A
FILING DATE: 19-DEC-95
ATTORNEY/AGENT INFORMATION:
NAME: Maniagouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: DFN-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2083 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 67..1390
US-08-574-959A-1

Query Match          0.9%; Score 18; DB 2; Length 2083;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 239 tgcacatggctctctc 256  
|||||  
Db 1271 tgcacatggctctctc 1288

## RESULT 11

US-07-593-657-8/c  
; Sequence 8, Application US/07593657  
; Patent No. 5266317  
; GENERAL INFORMATION:  
; APPLICANT: Tomalski, Michael D.  
; APPLICANT: Miller, Lois K.  
; TITLE OF INVENTION: INSECT-SPECIFIC PARALYTIC NEUROTOXIN  
; TITLE OF INVENTION: GENES FOR USE IN BIOLOGICAL INSECT CONTROL: METHODS AND  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee and Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: CO  
; COUNTRY: USA  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/593,657  
; FILING DATE: 19901004  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 14-90  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 303/499-8080  
; TELEFAX: 303/499-8089  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 155 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; US-07-593-657-8

Query Match 0.8%; Score 17; DB 1; Length 155;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1254 aaaaaaatacaatttt 1270  
|||||  
Db 32 AAAAAAATCAATTTT 16

RESULT 12  
US-07-593-657-9/c  
; Sequence 9, Application US/07593657  
; Patent No. 5266317  
; GENERAL INFORMATION:  
; APPLICANT: Tomalski, Michael D.

; APPLICANT: Miller, Lois K.  
; TITLE OF INVENTION: INSECT-SPECIFIC PARALYTIC NEUROTOXIN  
; TITLE OF INVENTION: GENES FOR USE IN BIOLOGICAL INSECT CONTROL: METHODS AND  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee and Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder

STATE: CO  
COUNTRY: USA  
ZIP: 80303

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/593,657  
FILING DATE: 19901004  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Ferber, Donna M.  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 14-90  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 303/499-8080  
TELEFAX: 303/499-8089

INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 156 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
US-07-593-657-9

Query Match 0.8%; Score 17; DB 1; Length 156;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1254 aaaaaaatacaatttt 1270  
|||||  
Db 52 AAAAAAATCAATTTT 36

RESULT 13  
US-07-593-657-11/c  
; Sequence 11, Application US/07593657  
; Patent No. 5266317  
; GENERAL INFORMATION:  
; APPLICANT: Tomalski, Michael D.  
; APPLICANT: Miller, Lois K.  
; TITLE OF INVENTION: INSECT-SPECIFIC PARALYTIC NEUROTOXIN  
; TITLE OF INVENTION: GENES FOR USE IN BIOLOGICAL INSECT CONTROL: METHODS AND  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee and Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: CO  
; COUNTRY: USA  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/593,657  
; FILING DATE: 19901004  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 14-90  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 303/499-8080  
; TELEFAX: 303/499-8089  
; INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:  
LENGTH: 163 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA to mRNA  
US-07-593-657-11

Query Match 0.8%; Score 17; DB 1; Length 163;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1254 aaaaaaatcaatttt 1270  
|||||  
DB 52 AAAAAAATCAATTTT 36

RESULT 14  
US-07-593-657-13/C  
Sequence 13 Application US/07593657  
Patent No. 5266317  
GENERAL INFORMATION:  
APPLICANT: Tomalski, Michael D.  
TITLE OF INVENTION: INSECT-SPECIFIC PARALYTIC NEUROTOXIN  
TITLE OF INVENTION: GENES FOR USE IN BIOLOGICAL INSECT CONTROL: METHODS AND  
TITLE OF INVENTION: COMPOSITIONS  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee and Winner, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: CO  
COUNTRY: USA  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/593,657  
FILING DATE: 19901004  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Felder, Donna M.  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 14-90  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 303/499-8080  
TELEFAX: 303/499-8089  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 210 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA to mRNA  
US-07-593-657-13

Query Match 0.8%; Score 17; DB 1; Length 210;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1254 aaaaaaatcaatttt 1270  
|||||  
DB 22 AAAAAAATCAATTTT 6

RESULT 15  
US-07-593-657-12/C

Sequence 12 Application US/07593657  
Patent No. 5266317  
GENERAL INFORMATION:  
APPLICANT: Tomalski, Michael D.  
TITLE OF INVENTION: INSECT-SPECIFIC PARALYTIC NEUROTOXIN  
TITLE OF INVENTION: GENES FOR USE IN BIOLOGICAL INSECT CONTROL: METHODS AND  
TITLE OF INVENTION: COMPOSITIONS  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee and Winner, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: CO  
COUNTRY: USA  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/593,657  
FILING DATE: 19901004  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Felder, Donna M.  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 14-90  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 303/499-8080  
TELEFAX: 303/499-8089  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 211 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA to mRNA  
US-07-593-657-12

Query Match 0.8%; Score 17; DB 1; Length 211;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1254 aaaaaaatcaatttt 1270  
|||||  
DB 22 AAAAAAATCAATTTT 6

RESULT 16  
US-07-593-657-14/C  
Sequence 14 Application US/07593657  
Patent No. 5266317  
GENERAL INFORMATION:  
APPLICANT: Tomalski, Michael D.  
TITLE OF INVENTION: INSECT-SPECIFIC PARALYTIC NEUROTOXIN  
TITLE OF INVENTION: GENES FOR USE IN BIOLOGICAL INSECT CONTROL: METHODS AND  
TITLE OF INVENTION: COMPOSITIONS  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee and Winner, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: CO  
COUNTRY: USA  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/593,657  
FILING DATE: 19901004  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Ferber, Donna M.  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 14-90  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 303/499-8080  
TELEFAX: 303/499-8089  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 319 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
US-07-593-657-14

Query Match 0.8%; Score 17; DB 1; Length 319;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1254 aaaaaaatcaatttt 1270  
Db 197 AAAAAATCAATT 181

RESULT 17  
US-09-276-531-79/C  
Sequence 79, Application US/09276531  
Patent No. 6183968  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Yue, Henry  
APPLICANT: Reddy, Koopa  
APPLICANT: Guegler, Karl J.  
APPLICANT: Baughn, Mariah R.  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING  
TITLE OF INVENTION: RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION  
NUMBER OF SEQUENCES: 134  
CORRESPONDENCE ADDRESS:  
ADDRESS: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: IBM PC compatible  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/276,531  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/079,677  
FILING DATE: March 27, 1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Lynn E. Murry, Ph.D.  
REGISTRATION NUMBER: 42,918  
REFERENCE/DOCKET NUMBER: PA-0008 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166

INFORMATION FOR SEQ ID NO: 79:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 510 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: COLONY06  
CLONE: 1610212  
US-09-276-531-79

Query Match 0.8%; Score 17; DB 4; Length 510;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1248 ttataaaaaaatca 1264  
Db 192 TTTATATAAAAAATCA 176

RESULT 18  
US-08-991-789A-42/C  
Sequence 42, Application US/08991789A  
Patent No. 6225054  
GENERAL INFORMATION:  
APPLICANT: Fridakis, Tony N.  
APPLICANT: Smith, John M.  
APPLICANT: Reed, Steven G.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
TREATMENT AND DIAGNOSIS OF BREAST CANCER  
NUMBER OF SEQUENCES: 292  
CORRESPONDENCE ADDRESS:  
ADDRESS: Seed IP Law Group  
STREET: 701 Fifth Avenue, Suite 6300  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/991,789A  
FILING DATE: 11-Dec-1997  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Potter, Jane E. R.  
REGISTRATION NUMBER: 33,332  
REFERENCE/DOCKET NUMBER: 210121.419C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 42:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 527 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 42:  
US-08-991-789A-42

Query Match 0.8%; Score 17; DB 4; Length 527;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1244 tacattataaaaaaa 1260  
Db 250 TACATTTATATAAAAA 234



RESULT 19  
US-08-905-223-19  
Sequence 19, Application US/08905223  
Patent No. 6222029  
GENERAL INFORMATION:  
APPLICANT: Edwards, Jean-Baptiste D.  
APPLICANT: Duclercq, Yveric  
APPLICANT: Lacroix, Bruno  
TITLE OF INVENTION: 5' ESTs FOR SECRETED PROTEINS  
NUMBER OF SEQUENCES: 503  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 501 West Broadway  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92101-3505  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Win95  
SOFTWARE: Word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/905,223  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelson, Ned A.  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 822 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: DOUBLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: CDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo Sapiens  
DEVELOPMENTAL STAGE: Fetal  
TISSUE TYPE: kidney  
FEATURE:  
NAME/KEY: other  
LOCATION: 260..464  
IDENTIFICATION METHOD: blastn  
OTHER INFORMATION: identity 96  
OTHER INFORMATION: region 153..357  
OTHER INFORMATION: id H57434  
OTHER INFORMATION: est  
FEATURE:  
NAME/KEY: other  
LOCATION: 118..184  
IDENTIFICATION METHOD: blastn  
OTHER INFORMATION: identity 98  
OTHER INFORMATION: region 98..164  
OTHER INFORMATION: id H57434  
OTHER INFORMATION: est  
FEATURE:  
NAME/KEY: other  
LOCATION: 56..113  
IDENTIFICATION METHOD: blastn  
OTHER INFORMATION: identity 98  
OTHER INFORMATION: region 35..92  
OTHER INFORMATION: id H57434  
OTHER INFORMATION: est  
FEATURE:  
NAME/KEY: other  
LOCATION: 454..485  
IDENTIFICATION METHOD: blastn

OTHER INFORMATION: identity 100  
OTHER INFORMATION: region 348..379  
OTHER INFORMATION: id H57434  
OTHER INFORMATION: est  
FEATURE:  
NAME/KEY: other  
LOCATION: 118..545  
IDENTIFICATION METHOD: blastn  
OTHER INFORMATION: identity 98  
OTHER INFORMATION: region 1..428  
OTHER INFORMATION: id N27248  
OTHER INFORMATION: est  
FEATURE:  
NAME/KEY: other  
LOCATION: 65..369  
IDENTIFICATION METHOD: blastn  
OTHER INFORMATION: identity 98  
OTHER INFORMATION: region 41..345  
OTHER INFORMATION: id H94779  
OTHER INFORMATION: est  
FEATURE:  
NAME/KEY: other  
LOCATION: 61..399  
IDENTIFICATION METHOD: blastn  
OTHER INFORMATION: identity 99  
OTHER INFORMATION: region 6..344  
OTHER INFORMATION: id H09880  
OTHER INFORMATION: est  
FEATURE:  
NAME/KEY: other  
LOCATION: 408..458  
IDENTIFICATION METHOD: blastn  
OTHER INFORMATION: identity 92  
OTHER INFORMATION: region 355..405  
OTHER INFORMATION: id H09880  
OTHER INFORMATION: est  
FEATURE:  
NAME/KEY: other  
LOCATION: 60..399  
IDENTIFICATION METHOD: blastn  
OTHER INFORMATION: identity 97  
OTHER INFORMATION: region 56..395  
OTHER INFORMATION: id H29351  
OTHER INFORMATION: est  
FEATURE:  
NAME/KEY: other  
LOCATION: 393..432  
IDENTIFICATION METHOD: blastn  
OTHER INFORMATION: identity 90  
OTHER INFORMATION: region 391..430  
OTHER INFORMATION: id H29351  
OTHER INFORMATION: est  
FEATURE:  
NAME/KEY: sig-peptide  
LOCATION: 346..408  
IDENTIFICATION METHOD: Von Heijne matrix  
OTHER INFORMATION: score 5.5  
OTHER INFORMATION: seq SFLPSALVIMTSA/AF

Query Match 0.8%; Score 17; DB 4; Length 822;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1887 ctttccccctgtgtgaa 1903  
|||||  
Db 733 CTTTCCCTGTGTAA 749

RESULT 20  
PCT-US93-05640-9/c  
Sequence 9, Application PC/TUS9305640

GENERAL INFORMATION:  
APPLICANT: University of Connecticut  
TITLE OF INVENTION: Contraceptive Vaccine  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/05640  
FILING DATE: 19930610  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/897,883  
FILING DATE: June 12, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Brook, David E.  
REGISTRATION NUMBER: 22,592  
REFERENCE/DOCKET NUMBER: UCT90-01AA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1104 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
PCT-US93-05640-9

Query Match 0.8%; Score 17; DB 5; Length 1104;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 420 tcttgcgaattccattc 436  
|||||  
DB 526 tcttgcgaattccattc 510

RESULT 21  
US-07-593-657-6/c  
Sequence 6, Application US/07593657  
Patent No. 5266317  
GENERAL INFORMATION:  
APPLICANT: Tomalski, Michael D.  
TITLE OF INVENTION: INSECT-SPECIFIC PARALYTIC NEUROTOXIN  
TITLE OF INVENTION: GENES FOR USE IN BIOLOGICAL INSECT CONTROL: METHODS AND  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee and Winner, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: CO  
COUNTRY: USA  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/593,657  
FILING DATE: 19901004  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Ferber, Donna M.  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 14-90  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 303/499-8080  
TELEFAX: 303/499-8089  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1241 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA to mRNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 119..985  
FEATURE:  
NAME/KEY: unsure  
LOCATION: -103..-100  
FEATURE:  
NAME/KEY: unsure  
LOCATION: 47..49  
US-07-593-657-6

Query Match 0.8%; Score 17; DB 1; Length 1241;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1254 aaaaaaatcaatttt 1270  
|||||  
DB 1158 AAAAAAATCAATTTT 1142

RESULT 22  
US-08-942-012B-3/c  
Sequence 3, Application US/08942012B  
Patent No. 6235278  
GENERAL INFORMATION:  
APPLICANT: Miller, Lois K.  
APPLICANT: Lu, Albert  
APPLICANT: Dierks, Peter  
APPLICANT: Black, Bruce  
TITLE OF INVENTION: Biological Insect Control  
TITLE OF INVENTION: Insect-Specific Toxin Genes,  
FILE REFERENCE: 28-96a  
CURRENT APPLICATION NUMBER: US/08/942,012B  
CURRENT FILING DATE: 1997-10-01  
PRIOR APPLICATION NUMBER: 08/729,606  
PRIOR FILING DATE: 2000-10-01  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: Patent Ver. 2.0  
SEQ ID NO 3  
LENGTH: 1241  
TYPE: DNA  
ORGANISM: Pycomotes tritici  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (119)..(985)  
US-08-942-012B-3

Query Match 0.8%; Score 17; DB 4; Length 1241;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1254 aaaaaaatcaatttt 1270  
|||||  
DB 1158 AAAAAAATCAATTTT 1142







Sequence 1, Application US/09062416  
Patent No. 6111094  
GENERAL INFORMATION:  
APPLICANT: C. Frank Bennett, Thomas P. Condon,  
APPLICANT: Shin Cheng Flounroy  
TITLE OF INVENTION: ENHANCED ANTISENSE MODULATION OF ICAM-1  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Offices of Jane Massey Licata  
STREET: 66 EAST MAIN STREET  
CITY: MARLTON  
STATE: NJ  
COUNTRY: USA  
ZIP: 08053  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION NUMBER: US/09/062,416  
APPLICATION NUMBER: 07/969,151  
FILING DATE: FEB 10, 1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 08/007,997  
FILING DATE: JAN 21, 1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/939,855  
FILING DATE: SEP 2, 1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/567,286  
FILING DATE: AUG 14, 1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Jane Massey Licata  
REGISTRATION NUMBER: 32,257  
REFERENCE/DOCKET NUMBER: ISPH-0306  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (609) 779-2400  
TELEFAX: (609) 810-1454  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2986 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Unknown  
MOLECULE TYPE: mRNA  
ANTI-SENSE: NO  
PUBLICATION INFORMATION:  
AUTHORS: Staunton D.E.  
AUTHORS: Marlin S.D.  
AUTHORS: Stracowa C.  
AUTHORS: Dustin M.L.  
AUTHORS: Springer T.A.  
TITLE: Primary structure of ICAM-1  
TITLE: demonstrates interaction between members of the  
TITLE: immunoglobulin and integrin supergene families  
JOURNAL: Cell  
VOLUME: 52  
ISSUE: 6  
PAGES: 925-933  
DATE: 25-MAR-1988  
US-09-062-416-1

Query Match 0.8%; Score 17; DB 3; Length 2986;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1253 aaaaaaaaaaaccattt 1269  
|||||  
Db 2911 AAAAAAAAAATCAATTT 2895

RESULT 31  
US-08-344-155C-97/C  
Sequence 97, Application US/08344155C  
Patent No. 5883082  
GENERAL INFORMATION:  
APPLICANT: Bennett and Stepkowski  
TITLE OF INVENTION: Compositions and Methods for Preventing  
TITLE OF INVENTION: and Treating Allograft Rejection  
NUMBER OF SEQUENCES: 99  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodland Falls Corporate Park  
STREET: 210 Lake Drive East, Suite 201  
CITY: Cherry Hill  
STATE: NJ  
COUNTRY: USA  
ZIP: 08002  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: WORDPERFECT 5.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/344,155C  
FILING DATE: No. 5883082ember 23, 1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 939,855  
FILING DATE: September 2, 1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/05209  
FILING DATE: July 23, 1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/063,167  
FILING DATE: 5/17/93  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/007,997  
FILING DATE: 1/21/93  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/939,855  
FILING DATE: 9/2/92  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/567,286  
FILING DATE: 8/14/90  
ATTORNEY/AGENT INFORMATION:  
NAME: Jane Massey Licata  
REGISTRATION NUMBER: 32,257  
REFERENCE/DOCKET NUMBER: ISPH-0098  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (609) 779-2400  
TELEFAX: (609) 779-8488  
INFORMATION FOR SEQ ID NO: 97:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3016  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
ANTI-SENSE: NO  
US-08-344-155C-97

Query Match 0.8%; Score 17; DB 2; Length 3016;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1253 aaaaaaaatcaattt 1269  
|||||  
Db 2924 AAAAAAAATCAATT 2908

## RESULT 32

5284931-1/c  
; Patent No. 5284931  
; APPLICANT: SPRINGER, TIMOTHY A.; ROTHLEIN, ROBERT; MARLIN,  
; STEVEN D.; DUSTIN, MICHAEL L.  
; TITLE OF INVENTION: INTERCELLULAR ADHESION MOLECULES AND  
; THEIR BINDING LIGANDS  
; NUMBER OF SEQUENCES: 41  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/515,478  
; FILING DATE: 27-APR-1990  
; SEQ ID NO: 1:  
; LENGTH: 3024  
5284931-1

Query Match 0.8%; Score 17; DB 6; Length 3024;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1253 aaaaaaaatcaattt 1269  
|||||  
Db 2911 AAAAAAAATCAATT 2895

## RESULT 33

US-08-366-577-1/c  
; Sequence 1, Application US/08366577  
; Patent No. 5728523  
; GENERAL INFORMATION:  
; APPLICANT: Vogelstein, Bert  
; APPLICANT: Kinzler, Kenneth W.  
; TITLE OF INVENTION: POLYMERASE DELTA MUTATIONS IN COLORECTAL  
; TITLES OF INVENTION: TUMORS WITH REPLICATION ERRORS  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner, Birch, McKie & Beckett  
; STREET: 1001 G Street N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20001

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/366,577  
FILING DATE: 12-DEC-1994  
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 01107.48554  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
TELEX: 197430 BBMB UT  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3435 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO

ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 43..3364  
US-08-366-577-1

Query Match 0.8%; Score 17; DB 1; Length 3435;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 42 gtgtccgcgcgcgcgt 58  
|||||  
Db 1379 GTGTCCGCGCGGCCGT 1363

## RESULT 34

PCT-US96-00005-1/c  
; Sequence 1, Application PC/TUS9600005  
; GENERAL INFORMATION:  
; APPLICANT: Vogelstein, Bert  
; APPLICANT: Kinzler, Kenneth W.  
; TITLE OF INVENTION: POLYMERASE DELTA MUTATIONS IN COLORECTAL  
; TITLES OF INVENTION: TUMORS WITH REPLICATION ERRORS  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner & Allegretti, Ltd.  
; STREET: 1001 G Street N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20001

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/00005  
FILING DATE: 2-JAN-96  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 01107.53505  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
TELEX: 197430 BBMB UT  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3435 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 43..3364  
PCT-US96-00005-1

Query Match 0.8%; Score 17; DB 5; Length 3435;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 42 gtgtccgcgcgcgcgt 58  
|||||

Db 1379 GTGTCCCGCCGCCGT 1363

## RESULT 35

US-08-696-834-5/C  
Sequence 5, Application US/08696834  
Patent No. 5834263  
GENERAL INFORMATION:  
APPLICANT: Niwa, Mineo  
APPLICANT: Saito, Yoshinasa  
APPLICANT: Ishii, Yoshinori  
APPLICANT: Yoshida, Masaru  
APPLICANT: Hayashi, Hiromi  
TITLE OF INVENTION: Method for Producing 2-Keto-L-Gulonic Acid  
NUMBER OF SEQUENCES: 48  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Oblon, Spivak, McCrelland, Maier & Neustadt,  
P.C.  
STREET: 1755 Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette - 3.50 inch, 1.44 Mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE:  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/696,834  
FILING DATE: 24-SEP-1996  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: JP 28612/1994  
FILING DATE: 25-FEB-1994  
ATTORNEY/AGENT INFORMATION:  
NAME:  
REGISTRATION NUMBER:  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4624 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Gluconobacter oxydans  
STRAIN: T-100  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1041..2534, 2537..4129  
IDENTIFICATION METHOD: experimentally  
FEATURE:  
NAME/KEY: mat peptide  
LOCATION: 1044..2534, 2540..4129  
IDENTIFICATION METHOD: experimentally  
US-08-696-834-5

Query Match 0.8%; Score 17; DB 2; Length 4624;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0y 41 cgtgtccgcgcgcgcgcg 57  
|||||  
Db 2659 CGTGTCCCGCCGCCCG 2643

## RESULT 36

US-09-077-354B-3  
Sequence 3, Application US/09077354B  
Patent No. 6255096  
GENERAL INFORMATION:  
APPLICANT: HOPWOOD, JOHN JOSEPH, SCOTT, HAMISH STEELE;  
APPLICANT: WEBER, BIRGIT; BLANCH, LIANNE; ANSON, DONALD STEWART  
TITLE OF INVENTION: SYNTHETIC MAMMALIAN  
TITLE OF INVENTION: N-ACETYLGLUCOSAMINIDASE AND GENETIC SEQUENCES ENCODING SA  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER  
STREET: 400 GARDEN CITY PLAZA  
CITY: GARDEN CITY  
STATE: NEW YORK  
COUNTRY: UNITED STATES  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/077,354B  
FILING DATE: 22-APRIL-1999  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/00747  
FILING DATE: 22-NOV-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: FOKALSKY, ANN R.  
REGISTRATION NUMBER: 34,697  
REFERENCE/DOCKET NUMBER: 12416  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 516 742 4343  
TELEFAX: 516 742 4366  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10380 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: Chromosome 17  
FEATURE:  
NAME/KEY: exon 1  
LOCATION: 990..1372  
FEATURE:  
NAME/KEY: exon 2  
LOCATION: 2115..2262  
FEATURE:  
NAME/KEY: exon 3  
LOCATION: 3056..3202  
FEATURE:  
NAME/KEY: exon 4  
LOCATION: 3387..3472  
FEATURE:  
NAME/KEY: exon 5  
LOCATION: 5667..5923  
FEATURE:  
NAME/KEY: exon 6  
LOCATION: 7745..8955  
US-09-077-354B-3

Query Match 0.8%; Score 17; DB 4; Length 10380;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0y 1327 gtaatccactggtgt 1343  
|||||



Db 6081 GTGATCCACTGTGT 6097

```
RESULT 37
US-08-670-175-2/C
; Sequence 2, Application US/08670175
; Patent No. 5854081
; GENERAL INFORMATION:
; APPLICANT: LINDEN, JOEL
; APPLICANT: TAYLOR, HEIDI
; APPLICANT: ROBEYA, ANNA
; APPLICANT: WOODARD, ROBIN
; APPLICANT: JIN, XIAMEI
; TITLE OF INVENTION: STABLE EXPRESSION OF HUMAN ADENOSINE
; TITLE OF INVENTION: RECEPTORS, AND ASSAYS EMPLOYING THE SAME
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ORLON, SPIYAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C. JEFFERSON DAVIS HIGHWAY, SUITE 400
; STREET: 1755 S.
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/670.175
; FILING DATE: 20-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ORLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 494-176-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 52 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "SYNTHETIC DNA PRIMER"
US-08-670-175-2

Query Match 0.8%; Score 16; DB 2; Length 52;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 322 gtcatgtatgtatga 337
Db 52 GTCATGTATGTATGA 37

RESULT 38
US-08-220-151-41
; Sequence 41, Application US/08220151
; Patent No. 5529780
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: Limbach, Keith J.
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF
; TITLE OF INVENTION: CANINE HERPESVIRUS gB, gC AND gD AND USES THEREFOR
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtiss, Morris & Safford
; STREET: 530 Fifth Avenue
```

```
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/220.151
; FILING DATE: 30-MAR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2540
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066 CURTMS
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 66 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
US-08-220-151-41

Query Match 0.8%; Score 16; DB 1; Length 66;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1865 taacaattttattt 1880
Db 13 TAAACAATTTTATT 28

RESULT 39
US-08-413-118-41
; Sequence 41, Application US/08413118
; Patent No. 5688920
; GENERAL INFORMATION:
; APPLICANT: PAOLETTI, ENZO
; APPLICANT: LIMBACH, KEITH J.
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF
; TITLE OF INVENTION: CANINE HERPESVIRUS gB, gC, AND gD AND USES THEREFOR
; NUMBER OF SEQUENCES: 128
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.
; STREET: 530 FIFTH AVENUE, 25TH FLOOR
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/413.118
; FILING DATE: 29-MAR-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/220.151
; FILING DATE: 30-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: FROMMER, WILLIAM S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2670
```

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 41:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 66 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-413-118-41

Query Match 0.8%; Score 16; DB 1; Length 66;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1865 taaacattttattt 1880  
DB 13 TAAACATTTTATT 28

RESULT 40  
US-08-224-657-18  
Sequence 18, Application US/08224657  
Patent No. 5756102  
GENERAL INFORMATION:  
APPLICANT: Paoletti, Enzo  
APPLICANT: Tartaglia, James  
APPLICANT: Taylor, Jill  
TITLE OF INVENTION: POXYVIRUS - CANINE DISTEMPER VIRUS (CDV)  
TITLE OF INVENTION: RECOMBINANTS AND COMPOSITIONS AND METHODS EMPLOYING THE  
NUMBER OF SEQUENCES: 122  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford, P.C.  
STREET: 530 Fifth Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/224,657  
FILING DATE: 06-APR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2550  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
TELEX: 425066 CURTMS  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 66 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-224-657-18

Query Match 0.8%; Score 16; DB 1; Length 66;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1865 taaacattttattt 1880

DB 13 TAAACATTTTATT 28

RESULT 41  
US-08-257-073-85  
Sequence 85, Application US/08257073  
Patent No. 5766597  
GENERAL INFORMATION:  
APPLICANT: Paoletti, Enzo  
APPLICANT: de Taisne, Charles  
APPLICANT: Tine, John A.  
TITLE OF INVENTION: MALARIA RECOMBINANT POXYVIRUS VACCINE  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford, P.C.  
STREET: 530 Fifth Avenue, 25th Floor  
CITY: New York  
STATE: New York  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/257,073  
FILING DATE: 09-JUN-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/075,783  
FILING DATE: 11-JUN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/852,305  
FILING DATE: 18-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/672,183  
FILING DATE: 20-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2570  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
TELEX: 425066 CURTMS  
INFORMATION FOR SEQ ID NO: 85:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 66 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-257-073-85

Query Match 0.8%; Score 16; DB 1; Length 66;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1865 taaacattttattt 1880  
DB 13 TAAACATTTTATT 28

RESULT 42  
US-08-184-009-18  
Sequence 18, Application US/08184009  
Patent No. 5833975  
GENERAL INFORMATION:  
APPLICANT: Paoletti, Enzo  
APPLICANT: Tartaglia, James  
APPLICANT: Cox, William I.

1  
TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY  
NUMBER OF SEQUENCES: 217  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CURTIS, MORRIS & SAFFORD  
STREET: 530 FIFTH AVENUE  
CITY: NEW YORK  
STATE: NY  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/184,009  
FILING DATE: 19-JAN-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: FROMMER, WILLIAM S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2530  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-0712  
TELEFAX: (212) 840-3333  
TELEX: 425066CURTMS  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 66 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-184-009-18

Query Match 0.8%; Score 16; DB 2; Length 66;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1865 taaacatttattt 1880  
|||||  
DB 13 TAAACATTTTATT 28

RESULT 43  
US-08-486-969-18  
Sequence 18, Application US/08486969  
Patent No. 3843456  
GENERAL INFORMATION:  
APPLICANT: PAOLETTI, ENZO  
APPLICANT: MAKI, JOANNE  
TITLE OF INVENTION: RECOMBINANT POXYVIRUS - RABIES  
TITLE OF INVENTION: COMPOSITIONS AND COMBINATION COMPOSITIONS AND USES  
NUMBER OF SEQUENCES: 55  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.  
STREET: 530 FIFTH AVENUE, 25th FLOOR  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: United States of America  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/486,969  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: FROMMER, WILLIAM S.

REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2600  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 66 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-486-969-18

Query Match 0.8%; Score 16; DB 2; Length 66;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1865 taaacatttattt 1880  
|||||  
DB 13 TAAACATTTTATT 28

RESULT 44  
US-08-417-210A-18  
Sequence 18, Application US/08417210A  
Patent No. 5863542  
GENERAL INFORMATION:  
APPLICANT: PAOLETTI, ENZO  
APPLICANT: TARTAGLIA, JAMES  
APPLICANT: COX, WILLIAM I.  
TITLE OF INVENTION: IMMUNODEFICIENCY RECOMBINANT POXYVIRUS  
NUMBER OF SEQUENCES: 148  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.  
STREET: 530 FIFTH AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/417,210A  
FILING DATE: 05-APR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: KOWALSKI, THOMAS J.  
REGISTRATION NUMBER: 32,147  
REFERENCE/DOCKET NUMBER: 454310-2690  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-840-3333  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 66 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-417-210A-18

Query Match 0.8%; Score 16; DB 2; Length 66;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1865 taaacatttattt 1880  
|||||  
DB 13 TAAACATTTTATT 28

RESULT 45  
US-08-458-356-18  
Sequence 18, Application US/08458356  
Patent No. 5942235  
GENERAL INFORMATION:  
APPLICANT: Paoletti, Enzo  
APPLICANT: Tartaglia, James  
APPLICANT: Cox, William I.  
TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY  
NUMBER OF SEQUENCES: 217  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford  
STREET: 530 Fifth Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/184,009  
FILING DATE: 19-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2530  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
TELEX: 425066CURTMS  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 66 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-458-356-18

Query Match 0.8%; Score 16; DB 2; Length 66;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1865 taaacattttattt 1880  
|||||  
DB 13 TAAACATTTTATT 28

RESULT 46  
US-08-471-025-18  
Sequence 18, Application US/08471025  
Patent No. 5989561  
GENERAL INFORMATION:  
APPLICANT: Paoletti, Enzo  
APPLICANT: Fischer, Laurent  
APPLICANT: Legros, Francois-Xavier  
TITLE OF INVENTION: RECOMBINANT POXVIRUS - CALICIVIRUS  
TITLE OF INVENTION: [RABBIT HEMORRHAGIC DISEASE VIRUS (RDHV)] COMPOSITIONS AND  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford, P.C.  
STREET: 530 Fifth Avenue, 25th Floor

CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: 06-JUN-1995  
APPLICATION NUMBER: US/08/471,025  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2650  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 66 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-471-025-18

Query Match 0.8%; Score 16; DB 2; Length 66;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1865 taaacattttattt 1880  
|||||  
DB 13 TAAACATTTTATT 28

RESULT 47  
US-08-473-446-41  
Sequence 41, Application US/08473446  
Patent No. 6017542  
GENERAL INFORMATION:  
APPLICANT: PAOLETTI, ENZO  
APPLICANT: LIMBACH, KEITH J.  
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF  
CANINE HERPESVIRUS 9B, 9C, AND 9D AND USES THEREFOR  
NUMBER OF SEQUENCES: 128  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.  
STREET: 530 FIFTH AVENUE, 25TH FLOOR  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: US/08/473,446  
APPLICATION NUMBER:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/413,118  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: FROMMER, WILLIAM S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2670  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 41:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 66 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-473-446-41

Query Match 0.8%; Score 16; DB 3; Length 66;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1865 taaacattttattt 1880  
DB 13 TAAACATTTTATT 28

RESULT 48  
PCT-US96-00547-18  
Sequence 18, Application PC/TUS9600547  
GENERAL INFORMATION:  
APPLICANT: Virogenetics Corporation  
TITLE OF INVENTION: RECOMBINANT POXVIRUS-HTLV, COMPOSITIONS  
TITLE OF INVENTION: AND USES  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford, P.C.  
STREET: 530 Fifth Avenue, 25th Floor  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/00547  
FILING DATE: 12-JAN-1996  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/372,664  
FILING DATE: 13-JAN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2621  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 66 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
PCT-US96-00547-18

Query Match 0.8%; Score 16; DB 5; Length 66;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1865 taaacattttattt 1880  
DB 13 TAAACATTTTATT 28

RESULT 49  
US-08-270-985-4  
Sequence 4, Application US/08270985  
Patent No. 5668012  
GENERAL INFORMATION:  
APPLICANT: Newman, Peter J.  
APPLICANT: Gumina, Richard J.  
APPLICANT: Kirschbaum, Nancy  
TITLE OF INVENTION: Platelet-Endothelial Cell Adhesion  
TITLE OF INVENTION: Molecule-1 Compositions and Methods  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: P. O. Box 1497  
CITY: Madison  
STATE: Wisconsin  
COUNTRY: USA  
ZIP: 53701-1497  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch,  
MEDIUM TYPE: 1.44 Mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: MS-DOS 3.3  
SOFTWARE: Wordperfect, Version 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/270,985  
FILING DATE: 5-July-94  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Scanlon, William J.  
REGISTRATION NUMBER: 30136  
REFERENCE/DOCKET NUMBER: 30383/137  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (608) 258-4284  
TELEFAX: (608) 258-4258  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 148 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-270-985-4

Query Match 0.8%; Score 16; DB 1; Length 148;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 442 ttattcttgagaa 457  
DB 106 TTATTTCTGAGAA 121

RESULT 50  
US-08-478-208-5  
Sequence 5, Application US/08478208A  
Patent No. 6087331  
GENERAL INFORMATION:  
APPLICANT: Newman, Peter J.  
APPLICANT: Kirschbaum, Nancy  
TITLE OF INVENTION: THERAPEUTIC USE OF PLATELET-ENDOTHELIAL CELL ADHESION  
TITLE OF INVENTION: MOLECULE-1 COMPOSITIONS  
FILE REFERENCE: 160180.90147  
CURRENT APPLICATION NUMBER: US/08/478,208A  
CURRENT FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: Patent Ver. 2.0  
SEQ ID NO 5  
LENGTH: 148

Mon Aug 20 10:21:42 2001

us-09-284-320-56.ol10.rni

Page 28

TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (20)..(127)  
US-08-478-208-5

Query Match 0.8%; Score 16; DB 3; Length 148;  
Best Local Similarity 100.0%; Pred. No. 2.le+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 442 ttatttctgagaaa 457  
Db 106 ttatttctgagaaa 121

Search completed: August 19, 2001, 00:21:20  
Job time: 9001 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2001, 01:09:16 ; Search time 5437.46 Seconds  
(without alignments)  
5783.203 Million cell updates/sec

Title: US-09-284-320-56  
Perfect score: 2033  
Sequence: 1 gactcgcagcgcgtcacctc.....gttaatgatgattcccaaac 2033

Scoring table:  
OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues

Word size : 10

Total number of hits satisfying chosen parameters: 1536093

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: listing first 1000 summaries

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_ba3:\*  
4: gb\_in1:\*  
5: gb\_in2:\*  
6: gb\_in3:\*  
7: gb\_om:\*  
8: gb\_ov:\*  
9: gb\_pat1:\*  
10: gb\_pat2:\*  
11: gb\_pi1:\*  
12: gb\_pi2:\*  
13: gb\_pi3:\*  
14: gb\_pi4:\*  
15: gb\_pi5:\*  
16: em\_ba1:\*  
17: em\_ba2:\*  
18: em\_fun:\*  
19: em\_hcgo\_hum:\*  
20: em\_hcgo\_inv:\*  
21: em\_hcgo\_rod:\*  
22: em\_hcgo\_hum1:\*  
23: em\_hcgo\_hum2:\*  
24: em\_hcgo\_hum3:\*  
25: em\_hcgo\_hum4:\*  
26: em\_hcgo\_hum5:\*  
27: em\_hcgo\_hum6:\*  
28: em\_hcgo\_hum7:\*  
29: em\_hcgo\_hum8:\*  
30: em\_hcgo\_inv1:\*  
31: em\_hcgo\_inv2:\*  
32: em\_hcgo\_rod:\*  
33: em\_hum1:\*  
34: em\_hum2:\*  
35: em\_hum3:\*  
36: em\_hum4:\*  
37: em\_hum5:\*  
38: em\_hum6:\*  
39: em\_hum7:\*  
40: em\_hum8:\*  
41: em\_in1:\*  
42: em\_om:\*  
43: em\_or:\*

44: em\_ov:\*  
45: em\_pat:\*  
46: em\_ph:\*  
47: em\_pi:\*  
48: em\_rod:\*  
49: em\_sts:\*  
50: em\_sy:\*  
51: em\_un:\*  
52: em\_v1:\*  
53: gb\_sts1:\*  
54: gb\_sts2:\*  
55: gb\_sts3:\*  
56: gb\_sy:\*  
57: gb\_un:\*  
58: gb\_v12:\*  
59: gb\_v12:\*  
60: gb\_hcg1:\*  
61: gb\_hcg2:\*  
62: gb\_hcg3:\*  
63: gb\_hcg4:\*  
64: gb\_hcg5:\*  
65: gb\_hcg6:\*  
66: gb\_hcg7:\*  
67: gb\_hcg8:\*  
68: gb\_hcg9:\*  
69: gb\_hcg10:\*  
70: gb\_hcg11:\*  
71: gb\_hcg12:\*  
72: gb\_hcg13:\*  
73: gb\_hcg14:\*  
74: gb\_hcg15:\*  
75: gb\_hcg16:\*  
76: gb\_hcg17:\*  
77: gb\_hcg18:\*  
78: gb\_hcg19:\*  
79: gb\_hcg20:\*  
80: gb\_hcg21:\*  
81: gb\_hcg22:\*  
82: gb\_hcg23:\*  
83: gb\_hcg24:\*  
84: gb\_hcg25:\*  
85: gb\_pi1:\*  
86: gb\_pi2:\*  
87: gb\_pi3:\*  
88: gb\_pi4:\*  
89: gb\_pi5:\*  
90: gb\_pi6:\*  
91: gb\_pi7:\*  
92: gb\_pi8:\*  
93: gb\_pi9:\*  
94: gb\_pi10:\*  
95: gb\_pi11:\*  
96: gb\_pi12:\*  
97: gb\_pi13:\*  
98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match length	DB ID	Description
1	1930	94.9	2049	AF248966
2	1776	87.4	1884	HSN800272
3	1078	53.0	12643	AC026156 Homo sapi
4	622	30.6	622	HSY17975
5	254	12.5	513	G27225
6	151	7.4	126543	AC026156 Homo sapi
7	116	5.7	148	A74462
8	116	5.7	148	A77441

9	116	5.7	151	9	A74460	A74460 Sequence 14	C	82	21	1.0	198692	78	AC090530
10	116	5.7	151	9	A74460	A74460 Sequence 14	C	83	21	1.0	227850	81	AL451143
11	116	5.7	204	9	A74463	A74463 Sequence 14	C	84	21	1.0	233941	83	CEY52B11
12	116	5.7	204	9	A74463	A74463 Sequence 14	C	85	21	1.0	283099	4	AE003466
13	78	3.8	158239	80	AL356315	AL356315 Homo sapi	C	86	20	1.0	344	54	G08563
14	78	3.8	165617	80	AL356381	AL356381 Homo sapi	C	87	20	1.0	345	54	G21686
15	78	3.8	165617	80	AL356381	AL356381 Homo sapi	C	88	20	1.0	382	92	HS191XD8
16	29	1.4	72831	94	AF259074	AF259074 Mus muscu	C	89	20	1.0	11515	90	AL391070
17	27	1.3	153519	79	AL354916	AL354916 Homo sapi	C	90	20	1.0	14043	1	AB014075
18	27	1.3	154776	81	AL512657	AL512657 Homo sapi	C	91	20	1.0	16070	6	CEY373A3A
19	23	1.1	162666	69	AC025817	AC025817 Homo sapi	C	92	20	1.0	16741	96	PFC030395
20	23	1.1	171093	68	AC026137	AC026137 Homo sapi	C	93	20	1.0	24685	65	AC020082
21	23	1.1	181413	68	AC023818	AC023818 Homo sapi	C	94	20	1.0	30544	90	AL139108
22	22	1.1	11230	5	CEAH10	AL023822 Caenorhab	C	95	20	1.0	36532	91	AP001223
23	22	1.1	31039	5	CEAH10	281027 Caenorhab	C	96	20	1.0	43410	5	HSU242F8
24	22	1.1	41524	4	AC006794	AC006794 Caenorhab	C	97	20	1.0	43410	5	HSU242F8
25	22	1.1	43236	6	CELC33H5	U41007 Caenorhabd	C	98	20	1.0	43410	5	HSU242F8
26	22	1.1	107183	12	CEY102A5A	AC002983 Arabidops	C	99	20	1.0	47336	4	AC004353
27	22	1.1	110000	83	CEY102A5A_0	299711 Caenorhabd	C	100	20	1.0	47827	13	AP000731
28	22	1.1	128468	90	AL139090	AL139090 Human DNA	C	101	20	1.0	58069	4	AC024775
29	22	1.1	167366	66	AC021774	AC021774 Homo sapi	C	102	20	1.0	58977	60	AC008251
30	22	1.1	171185	83	AP002424	AP002424 Homo sapi	C	103	20	1.0	61864	6	CEY50E8A
31	22	1.1	177097	82	AP001569	AP001569 Homo sapi	C	104	20	1.0	79054	14	F21F23
32	22	1.1	179726	86	AC007052	AC007052 Homo sapi	C	105	20	1.0	83557	89	AL138735
33	22	1.1	182411	78	AC090408	AC090408 Homo sapi	C	106	20	1.0	94639	13	ATP1AP22
34	22	1.1	196869	77	AC087535	AC087535 Homo sapi	C	107	20	1.0	95937	12	ATP23B13
35	22	1.1	199362	13	ATC8RTV24	AL161512 Arabidops	C	108	20	1.0	109180	12	AC064879
36	22	1.1	200774	82	AP001592	AP001592 Arabidops	C	109	20	1.0	110000	83	CEY38E10_2
37	22	1.1	326419	76	AC079633	AC079633 Mus muscu	C	110	20	1.0	110932	6	CEY38E10A
38	22	1.0	2061	89	AK021970	AK021970 Homo sapi	C	111	20	1.0	111330	96	FSMAL3P3
39	22	1.0	27535	12	AC006224	AC006224 Arabidops	C	112	20	1.0	114736	93	HS100419
40	22	1.0	43784	64	AC017442	AC017442 Drosophill	C	113	20	1.0	117757	93	CEY38E10A
41	22	1.0	90130	90	AL161618	AL161618 Human DNA	C	114	20	1.0	118264	85	AB045358
42	22	1.0	101798	81	AL390960	AL390960 Homo sapi	C	115	20	1.0	120920	61	AC010398
43	22	1.0	103443	12	AC010164	AC010164 Arabidops	C	116	20	1.0	121668	12	AC022521
44	22	1.0	106949	6	CEY52B11A	AL022654 Caenorhab	C	117	20	1.0	122014	90	AL139833
45	22	1.0	116861	60	AC007985	AC007985 Drosophill	C	118	20	1.0	125235	87	AC008842
46	22	1.0	116844	12	AC005142	AF289077 Arabidops	C	119	20	1.0	127936	67	AC022099
47	22	1.0	145174	89	AF289077	AF289077 Homo sapi	C	120	20	1.0	129779	90	AL359457
48	22	1.0	150880	78	AL362038	AL362038 Homo sapi	C	121	20	1.0	131025	78	AL139162
49	22	1.0	151320	72	AL136079	AL136079 Homo sapi	C	122	20	1.0	132466	63	AC015608
50	22	1.0	153266	67	AC023306	AC023306 Homo sapi	C	123	20	1.0	134931	87	AC010244
51	22	1.0	154716	74	AC069502	AC069502 Homo sapi	C	124	20	1.0	138550	60	AC006721
52	22	1.0	156165	66	AC021453	AC021453 Homo sapi	C	125	20	1.0	139714	68	AC024322
53	22	1.0	159629	13	ATC8RTV9	AL161497 Arabidops	C	126	20	1.0	140469	69	AC025066
54	22	1.0	160940	81	AL512299	AL512299 Homo sapi	C	127	20	1.0	143220	90	AL389887
55	22	1.0	163659	82	AP001802	AP001802 Homo sapi	C	128	20	1.0	143536	66	AC020846
56	22	1.0	166000	80	AL390731	AL390731 Human DNA	C	129	20	1.0	144778	84	CNS07E8F
57	22	1.0	168142	80	AL356245	AL356245 Homo sapi	C	130	20	1.0	147620	62	AC011195
58	22	1.0	168821	28	AL3591835	AL3591835 Human DNA	C	131	20	1.0	149064	61	AC009445
59	22	1.0	169434	90	AL359205	AL359205 Homo sapi	C	132	20	1.0	152142	91	AP003494
60	22	1.0	171526	69	AC026262	AC026262 Homo sapi	C	133	20	1.0	153339	82	AL590635
61	22	1.0	171588	69	AC023198	AC023198 Homo sapi	C	134	20	1.0	156612	72	AC053515
62	22	1.0	172058	78	AP188028	AP188028 Homo sapi	C	135	20	1.0	153084	76	AC079888
63	22	1.0	172966	71	AC009707	AC009707 Homo sapi	C	136	20	1.0	154212	90	AL354829
64	22	1.0	174143	86	AC007486	AC007486 Homo sapi	C	137	20	1.0	154419	78	AF165139
65	22	1.0	174725	85	AC0018915	AC0018915 Homo sapi	C	138	20	1.0	154814	67	AC022934
66	22	1.0	175054	86	AC0018915	AL161432 Homo sapi	C	139	20	1.0	156182	65	AC019031
67	22	1.0	176762	79	AP002853	AP002853 Homo sapi	C	140	20	1.0	156923	66	AC021256
68	22	1.0	177464	91	AL161432	AL161432 Homo sapi	C	141	20	1.0	157268	69	AC025063
69	22	1.0	182035	76	AP127019	AP127019 Homo sapi	C	142	20	1.0	159324	81	AL512447
70	22	1.0	182505	78	AP127019	AL354711 Homo sapi	C	143	20	1.0	161869	90	AL390882
71	22	1.0	182905	67	AC022885	AC022885 Homo sapi	C	144	20	1.0	162066	79	AL354711
72	22	1.0	183746	80	AL389921	AL389921 Homo sapi	C	145	20	1.0	162299	75	AC027121
73	22	1.0	184232	80	AC021535	AC021535 Homo sapi	C	146	20	1.0	163662	90	AL390882
74	22	1.0	184466	82	AC012113	AC012113 Homo sapi	C	147	20	1.0	163771	88	AC027121
75	22	1.0	184663	80	AC012113	AC012113 Homo sapi	C	148	20	1.0	163771	88	AC027121
76	22	1.0	184917	75	AC073732	AC073732 Mus muscu	C	149	20	1.0	163771	88	AC027121
77	22	1.0	185614	76	AC008104	AC008104 Homo sapi	C	150	20	1.0	163771	88	AC027121
78	22	1.0	185614	76	AC009466	AC009466 Homo sapi	C	151	20	1.0	163771	88	AC027121
79	22	1.0	185729	87	AC009466	AC009466 Homo sapi	C	152	20	1.0	163771	88	AC027121
80	22	1.0	193414	74	AC073696	AC073696 Mus muscu	C	153	20	1.0	165142	70	AC026864
81	22	1.0	193414	74	AC073696	AC073696 Mus muscu	C	154	20	1.0	166338	88	AC026165



C 155	20	1.0	167164	70	AC026862	Homo sapi	C 228	19	0.9	7900	14	AY009938
C 156	20	1.0	167201	90	AL392044	Human DNA	C 229	19	0.9	8074	14	U0302293
C 157	20	1.0	168311	71	AC044841	Homo sapi	C 230	19	0.9	8535	14	MTSMGNS
C 158	20	1.0	168311	93	HSBA101J8	Human DNA	C 231	19	0.9	12720	1	AE005453
C 159	20	1.0	168843	79	AL157772	Homo sapi	C 232	19	0.9	13275	1	AE000310
C 160	20	1.0	168909	87	AC010165	Homo sapi	C 233	19	0.9	19791	2	D90850
C 161	20	1.0	169786	74	AC073416	Homo sapi	C 234	19	0.9	20505	5	CBRG18K06
C 162	20	1.0	170333	78	AL139019	Homo sapi	C 235	19	0.9	24784	6	CELP29B9
C 163	20	1.0	172476	86	AC007736	Homo sapi	C 236	19	0.9	25480	63	AC012841
C 164	20	1.0	172601	66	AC021455	Homo sapi	C 237	19	0.9	25818	63	AC014479
C 165	20	1.0	172758	67	AC022553	Homo sapi	C 238	19	0.9	26624	12	AB011481
C 166	20	1.0	174274	71	AC044505	Homo sapi	C 239	19	0.9	27533	6	CELP28B4
C 167	20	1.0	175067	71	AC040898	Homo sapi	C 240	19	0.9	27848	6	CEM01E5
C 168	20	1.0	175954	79	AL354920	Homo sapi	C 241	19	0.9	30152	6	CELP19C3
C 169	20	1.0	177933	91	AP002759	Homo sapi	C 242	19	0.9	34087	4	CEH02112
C 170	20	1.0	178213	83	AP003468	Homo sapi	C 243	19	0.9	35030	4	AC024794
C 171	20	1.0	179367	79	AL355874	Homo sapi	C 244	19	0.9	35030	4	AC024794
C 172	20	1.0	180864	69	AC025846	Homo sapi	C 245	19	0.9	36267	6	CELP02C5
C 173	20	1.0	180977	66	AC020524	Homo sapi	C 246	19	0.9	37385	6	CELP29C4
C 174	20	1.0	182049	66	AC021239	Homo sapi	C 247	19	0.9	38731	6	CELP32A5
C 175	20	1.0	183019	70	AC027564	Homo sapi	C 248	19	0.9	39589	6	CELP32A5
C 176	20	1.0	185621	71	AC036195	Homo sapi	C 249	19	0.9	40280	5	CBRG17N15
C 177	20	1.0	187892	82	AL589844	Homo sapi	C 250	19	0.9	40641	4	AC006816
C 178	20	1.0	190119	82	AP001814	Homo sapi	C 251	19	0.9	40901	6	CELP14D12
C 179	20	1.0	190211	88	AC025770	Homo sapi	C 252	19	0.9	43046	6	CELP09B8
C 180	20	1.0	190515	83	CNS01DX4	Homo sapi	C 253	19	0.9	43637	5	CEBP399
C 181	20	1.0	190802	70	AC007448	Homo sapi	C 254	19	0.9	45510	5	CEC32A3
C 182	20	1.0	191923	60	AC026958	Homo sapi	C 255	19	0.9	47269	9	AX059551
C 183	20	1.0	192518	84	HSBA539A6	Homo sapi	C 256	19	0.9	49448	9	AX059542
C 184	20	1.0	194457	82	AL590076	Homo sapi	C 257	19	0.9	55432	64	AC017613
C 185	20	1.0	194591	61	AC009677	Homo sapi	C 258	19	0.9	57493	85	AC004505
C 186	20	1.0	194750	61	AC009677	Homo sapi	C 259	19	0.9	64998	63	AC014035
C 187	20	1.0	199840	88	AC077690	Homo sapi	C 260	19	0.9	65239	65	AC019844
C 188	20	1.0	204018	64	AC016672	Homo sapi	C 261	19	0.9	66107	68	AC023840
C 189	20	1.0	207684	65	AC018039	Homo sapi	C 262	19	0.9	66881	67	AC023178
C 190	20	1.0	208622	78	AL138705	Homo sapi	C 263	19	0.9	67560	78	AC090848
C 191	20	1.0	215234	75	AC073745	Mus muscu	C 264	19	0.9	67780	68	AC023690
C 192	20	1.0	217384	68	AC023573	Homo sapi	C 265	19	0.9	68279	77	AC090165
C 193	20	1.0	219888	76	AC079959	Mus muscu	C 266	19	0.9	72907	64	AC016010
C 194	20	1.0	222400	61	AC094440	Homo sapi	C 267	19	0.9	73851	90	AL356464
C 195	20	1.0	223020	90	AL162424	Human DNA	C 268	19	0.9	75522	92	HS116845
C 196	20	1.0	224583	75	AC074153	Mus muscu	C 269	19	0.9	76065	66	AC031250
C 197	20	1.0	226595	72	AC060782	Mus muscu	C 270	19	0.9	76738	89	AL138813
C 198	20	1.0	226909	76	AC079490	Mus muscu	C 271	19	0.9	76738	89	AC014256
C 199	20	1.0	263693	5	AE003793	Drosophila	C 272	19	0.9	76748	63	AC014256
C 200	20	1.0	285495	4	AE003560	Drosophila	C 273	19	0.9	77378	64	AC016183
C 201	20	1.0	300000	91	AP002532	Homo sapi	C 274	19	0.9	78276	6	CELY38C1AA
C 202	19	0.9	129	2	ECMICE	X04888 E. coli micf	C 275	19	0.9	81207	92	HS12513
C 203	19	0.9	174	10	E00507	E00507 micf mRNA s	C 276	19	0.9	81419	89	AL133382
C 204	19	0.9	358	3	S75172	CMPC-outer	C 277	19	0.9	82073	88	AC020979
C 205	19	0.9	371	10	AX100261	Sequence	C 278	19	0.9	83550	89	AF312912
C 206	19	0.9	854	53	CNS074AD	AL28879 clone BAO	C 279	19	0.9	83693	13	AT11338
C 207	19	0.9	855	53	CNS074AD	AL28879 clone BAO	C 280	19	0.9	84203	12	AC005106
C 208	19	0.9	870	96	OCMP	Y1351 Osterlagia	C 281	19	0.9	84896	12	AB025639
C 209	19	0.9	898	94	AF201698	AF201698 Mus muscu	C 282	19	0.9	85233	6	CELY55D5A
C 210	19	0.9	929	53	CNS06Y82	AL20696 T3 end of	C 283	19	0.9	91640	92	HS426F10
C 211	19	0.9	984	53	CNS06Y82	AL30909 clone XBA	C 284	19	0.9	91820	6	CELY92H12A
C 212	19	0.9	1001	53	CNS06Y82	AL22578 clone BAO	C 285	19	0.9	97433	81	AL512789
C 213	19	0.9	1027	53	CNS07160	AB025275 Stichopus	C 286	19	0.9	100000	91	AP000075
C 214	19	0.9	1439	4	AB025275	AB025275 Stichopus	C 287	19	0.9	101640	6	CELY54E10A
C 215	19	0.9	1545	3	KPOMP366	233506 K. pneumoniae	C 288	19	0.9	103970	6	CELY55B8A
C 216	19	0.9	1629	1	AF039309	M31424 S. typhi out	C 289	19	0.9	104148	86	AC006146
C 217	19	0.9	1639	3	STYOMPC	273184 S. typhi out	C 290	19	0.9	104785	91	AP001435
C 218	19	0.9	1660	15	SCYLR012C	273184 S. typhi out	C 291	19	0.9	105223	12	AC007399
C 219	19	0.9	1713	2	ECOMPC	273184 S. typhi out	C 292	19	0.9	106702	13	ATP7K2
C 220	19	0.9	1713	2	ECOMPC	273184 S. typhi out	C 293	19	0.9	107102	85	AC004946
C 221	19	0.9	1907	14	NTP32	273184 S. typhi out	C 294	19	0.9	107885	86	AC006389
C 222	19	0.9	1917	95	SCYLR013W	273184 S. typhi out	C 295	19	0.9	108090	87	AC009553
C 223	19	0.9	1917	95	SCYLR013W	273184 S. typhi out	C 296	19	0.9	108460	92	HS111C20
C 224	19	0.9	2111	3	SMOMPC	273184 S. typhi out	C 297	19	0.9	109694	14	F23A5
C 225	19	0.9	2308	95	RATKINIA	273184 S. typhi out	C 298	19	0.9	110000	28	AP002753-2
C 226	19	0.9	2386	95	RATKINIC	273184 S. typhi out	C 299	19	0.9	110000	83	CEV116F11-3
C 227	19	0.9	3332	8	GGINTA6	X56559 Chicken MRN	C 300	19	0.9	110000	83	CEV116F11-4

AY009938 Hordeum v  
 A3302293 Hordeum v  
 Z47795 P. subcordif  
 AE005453 Escherich  
 AE000310 Escherich  
 D90850 E. coli geno  
 AC084519 Caenorhab  
 U70849 Caenorhabd  
 AC012841 Drosophila  
 AC014479 Drosophila  
 AB011481 Arabidops  
 AF026206 Caenorhab  
 Z93385 Caenorhabd  
 U28412 Caenorhabd  
 Z92789 Caenorhabd  
 AC024794 Caenorhab  
 AC024794 Caenorhab  
 U55374 Caenorhabd  
 AF067616 Caenorhab  
 U20864 Caenorhabd  
 AF038605 Caenorhab  
 AC084515 Caenorhab  
 AC006816 Caenorhab  
 U41021 Caenorhabd  
 U29612 Caenorhabd  
 Z81455 Caenorhabd  
 Z48241 Caenorhabd  
 Z48241 Caenorhabd  
 AX059551 Sequence  
 AX059542 Sequence  
 AC017613 Drosophila  
 AC004505 Homo sapi  
 AC014035 Drosophila  
 AC019844 Drosophila  
 AC023840 Homo sapi  
 AC023178 Homo sapi  
 AC090848 Homo sapi  
 AC023690 Drosophila  
 AC090165 Homo sapi  
 AC016010 Homo sapi  
 AL356464 Human DNA  
 AL031183 Human DNA  
 AC006921 Arabidops  
 AC021250 Homo sapi  
 AL138813 Human DNA  
 AC014256 Drosophila  
 AC016183 Homo sapi  
 AC024761 Caenorhab  
 AL033528 Human DNA  
 AL133382 Human DNA  
 AC020979 Homo sapi  
 AP312912 Homo sapi  
 AL035524 Arabidops  
 AC005106 Genomic s  
 AB025639 Arabidops  
 AC084196 Caenorhab  
 AL023358 Caenorhab  
 AC084161 Caenorhab  
 AL512789 Homo sapi  
 AP000075 Homo sapi  
 AC024810 Caenorhab  
 AC024877 Caenorhab  
 AC006146 Homo sapi  
 AP001435 Homo sapi  
 AC007399 Arabidops  
 AL033545 Arabidops  
 AC004946 Homo sapi  
 AC006389 Homo sapi  
 AC009553 Homo sapi  
 AL035530 Human DNA  
 AC011713 Arabidops  
 Continuation (3 of  
 Continuation (4 of  
 Continuation (5 of

301	19	0.9	110000	83	CEY43F8_2	Continuation (3 of	374	19	0.9	156210	64	AC016303	AC016303 Homo sapi
302	19	0.9	110000	83	CEY43F8_3	Continuation (4 of	375	19	0.9	156486	62	AC011657	AC011657 Homo sapi
303	19	0.9	110079	94	AC002315	AC002315 Mouse BAC	376	19	0.9	156563	76	AC079458	AC079458 Homo sapi
304	19	0.9	112629	14	F2806	AF147762 Arabidops	377	19	0.9	156754	66	AC021364	AC021364 Homo sapi
305	19	0.9	114200	81	AL391556	AL391536 Homo sapi	378	19	0.9	156906	64	AC016295	AC016295 Homo sapi
306	19	0.9	115202	62	AC012310	AC012310 Homo sapi	379	19	0.9	157067	12	AC007627	AC007627 Genomic S
307	19	0.9	116457	89	AL137845	AL137845 Human DNA	380	19	0.9	157067	87	AC010163	AC010163 Homo sapi
308	19	0.9	120407	83	AP001967	AP001967 Homo sapi	381	19	0.9	157141	87	AC011470	AC011470 Homo sapi
309	19	0.9	120511	75	AC079266	AC079266 Mus muscu	382	19	0.9	158022	90	AL445306	AL445306 Homo sapi
310	19	0.9	120950	6	CEY5007A	AC024801 Caenorhab	383	19	0.9	158037	62	AC012126	AC012126 Homo sapi
311	19	0.9	121054	13	ATAC013483	AC031483 Arabidops	384	19	0.9	158078	90	AL391821	AL391821 Homo sapi
312	19	0.9	121703	86	AC006452	AC006452 Homo sapi	385	19	0.9	159284	70	AC084369	AC084369 Homo sapi
313	19	0.9	122568	89	AF196972	AF196972 Homo sapi	386	19	0.9	159634	81	AL391425	AL391425 Homo sapi
314	19	0.9	122855	6	CELY383A	AC084159 Caenorhab	387	19	0.9	159716	69	AC025807	AC025807 Homo sapi
315	19	0.9	123010	88	AC087321	AC087321 Homo sapi	388	19	0.9	160690	98	AC023457	AC023457 Homo sapi
316	19	0.9	126140	4	AC007329	AC007329 Drosophill	389	19	0.9	160743	93	HS256F22	HS256F22 Homo sapi
317	19	0.9	127247	92	HS326L13	282170 Human DNA s	390	19	0.9	161178	62	AC011178	AC011178 Homo sapi
318	19	0.9	128559	85	AC004677	AC004677 Homo Sapi	391	19	0.9	161886	76	AC079595	AC079595 Homo sapi
319	19	0.9	128769	87	AC013357	AC013357 Homo sapi	392	19	0.9	162078	60	AC008661	AC008661 Homo sapi
320	19	0.9	129060	6	CEY41CA4	AL032627 Caenorhab	393	19	0.9	162099	63	AC015730	AC015730 Homo sapi
321	19	0.9	129467	93	HS1601K24	AL109626 Human DNA	394	19	0.9	162354	89	AF260225	AF260225 Homo sapi
322	19	0.9	131684	61	AC010577	AC010577 Drosophill	395	19	0.9	162487	76	AC079947	AC079947 Homo sapi
323	19	0.9	132860	62	AC011068	AC011068 Drosophill	396	19	0.9	162589	82	AL590068	AL590068 Homo sapi
324	19	0.9	132885	83	AP003412	AP003412 Oryza sat	397	19	0.9	162595	4	AC008335	AC008335 Drosophill
325	19	0.9	134095	60	AC006915	AC006915 Caenorhab	398	19	0.9	162868	87	AC010348	AC010348 Homo sapi
326	19	0.9	136810	62	AC012172	AC012172 Homo sapi	399	19	0.9	162801	75	AC079137	AC079137 Homo sapi
327	19	0.9	137521	84	CNS06C82	AL391156 Homo sapi	400	19	0.9	163022	67	AC022756	AC022756 Homo sapi
328	19	0.9	137945	78	AL133508	AL133508 Homo sapi	401	19	0.9	163386	88	AC021699	AC021699 Homo sapi
329	19	0.9	138621	91	AP002028	AP002028 Homo sapi	402	19	0.9	163495	76	AC024183	AC024183 Homo sapi
330	19	0.9	139291	6	CELY82E9BR	AC090929 Caenorhab	403	19	0.9	163563	78	AC079443	AC079443 Mus muscu
331	19	0.9	139447	81	AL1513220	AL1513220 Homo sapi	404	19	0.9	164083	71	AC040969	AC040969 Homo sapi
332	19	0.9	140765	88	AC021301	AC021301 Homo sapi	405	19	0.9	164232	61	AC010855	AC010855 Homo sapi
333	19	0.9	143893	75	AC079139	AC079139 Homo sapi	406	19	0.9	164511	60	AC008110	AC008110 Homo sapi
334	19	0.9	144260	86	AC006998	AC006998 Homo sapi	407	19	0.9	164801	68	AC023411	AC023411 Homo sapi
335	19	0.9	144745	72	AC060827	AC060827 Homo sapi	408	19	0.9	164891	57	AC022848	AC022848 Homo sapi
336	19	0.9	145166	72	AC053473	AC053473 Homo sapi	409	19	0.9	164908	76	AC080009	AC080009 Homo sapi
337	19	0.9	145417	62	AC012303	AC012303 Homo sapi	410	19	0.9	164985	62	AC012439	AC012439 Homo sapi
338	19	0.9	145453	12	AC012477	AC012477 Genomic S	411	19	0.9	165038	70	AC027401	AC027401 Homo sapi
339	19	0.9	145614	60	AC006872	AC006872 Caenorhab	412	19	0.9	165524	72	AC067068	AC067068 Homo sapi
340	19	0.9	145687	63	AC013646	AC013646 Homo sapi	413	19	0.9	165562	73	AC068086	AC068086 Homo sapi
341	19	0.9	145772	61	AC010467	AC010467 Homo sapi	414	19	0.9	165654	64	AC016762	AC016762 Homo sapi
342	19	0.9	145939	79	AL158143	AL158143 Homo sapi	415	19	0.9	166032	68	AC023645	AC023645 Homo sapi
343	19	0.9	146688	88	AC022261	AC022261 Homo sapi	416	19	0.9	166122	90	AL157815	AL157815 Homo sapi
344	19	0.9	146797	60	AC008284	AC008284 Drosophill	417	19	0.9	166417	64	AC016964	AC016964 Homo sapi
345	19	0.9	147784	89	AL136371	AL136371 Human DNA	418	19	0.9	166465	84	CNS07EFW	AL583743 Homo sapi
346	19	0.9	148005	64	AC024444	AC024444 Homo sapi	419	19	0.9	166522	80	AL357072	AL357072 Homo sapi
347	19	0.9	148245	69	AC016209	AC016209 Homo sapi	420	19	0.9	166971	90	AL157812	AL157812 Human DNA
348	19	0.9	148352	69	AC025309	AC025309 Homo sapi	421	19	0.9	166993	72	AC064877	AC064877 Homo sapi
349	19	0.9	149749	90	AL445236	AL445236 Human DNA	422	19	0.9	167509	62	AC012107	AC012107 Homo sapi
350	19	0.9	149797	80	AL456797	AL456797 Homo sapi	423	19	0.9	167961	83	AP002422	AP002422 Homo sapi
351	19	0.9	150424	90	AL391811	AL391811 Human DNA	424	19	0.9	168651	75	AC074363	AC074363 Homo sapi
352	19	0.9	150759	82	AP001130	AP001130 Homo sapi	425	19	0.9	168701	71	AC013573	AC013573 Homo sapi
353	19	0.9	150785	69	AC026158	AC026158 Homo sapi	426	19	0.9	168741	63	AC015753	AC015753 Homo sapi
354	19	0.9	150799	68	AC024402	AC024402 Homo sapi	427	19	0.9	168986	86	AC008013	AC008013 Homo sapi
355	19	0.9	151040	66	AC020827	AC020827 Mus muscu	428	19	0.9	169101	74	AC008317	AC008317 Drosophill
356	19	0.9	151770	93	HSAC002066	AC002066 Human BAC	429	19	0.9	169172	74	AC069537	AC069537 Homo sapi
357	19	0.9	151979	69	AC025113	AC025113 Homo sapi	430	19	0.9	169181	67	AC023303	AC023303 Homo sapi
358	19	0.9	152369	84	CNS07ED2	AL445443 Homo sapi	431	19	0.9	169226	60	AC006913	AC006913 Caenorhab
359	19	0.9	153006	71	AC027782	AC027782 Homo sapi	432	19	0.9	169388	69	AC005243	AC005243 Homo sapi
360	19	0.9	153133	73	AC068405	AC068405 Homo sapi	433	19	0.9	169543	70	AC025826	AC025826 Homo sapi
361	19	0.9	153615	70	AC027168	AC027168 Homo sapi	434	19	0.9	169597	87	AC009299	AC009299 Homo sapi
362	19	0.9	154243	72	AC055768	AC055768 Homo sapi	435	19	0.9	169758	80	AL356139	AL356139 Homo sapi
363	19	0.9	154470	74	AC006982	AC006982 Homo sapi	436	19	0.9	169807	81	AL512665	AL512665 Homo sapi
364	19	0.9	154959	69	AC025125	AC025125 Homo sapi	437	19	0.9	170000	60	AC004524	AC004524 Homo sapi
365	19	0.9	155285	86	AC025939	AC025939 Homo sapi	438	19	0.9	170216	66	AC021701	AC021701 Homo sapi
366	19	0.9	155450	86	AC005951	AC005951 Homo sapi	439	19	0.9	170640	72	AC048375	AC048375 Homo sapi
367	19	0.9	155703	73	AC067959	AC067959 Homo sapi	440	19	0.9	171129	86	AC006031	AC006031 Homo sapi
368	19	0.9	155757	76	AC079984	AC079984 Homo sapi	441	19	0.9	171235	63	AC013595	AC013595 Homo sapi
369	19	0.9	155860	87	AC009085	AC009085 Homo sapi	442	19	0.9	171715	82	AP001148	AP001148 Homo sapi
370	19	0.9	155939	14	CHNTXX	Z00044 Nicotiana t	443	19	0.9	171735	84	AP001652	AP001652 Homo sapi
371	19	0.9	156075	90	AL358855	AL358855 Human DNA	444	19	0.9	171495	82	AL137069	AL137069 Human DNA
372	19	0.9	156102	64	AC016081	AC016081 Homo sapi	445	19	0.9	171852	89	AL137069	AL137069 Human DNA
373	19	0.9	156155	74	AC073120	AC073120 Homo sapi	446	19	0.9	171852	89	AL137069	AL137069 Human DNA

C 447	19	0.9 171917	77	AC084836	C 520	19	0.9 196720	91	AP002026
C 448	19	0.9 171945	69	AC023557	C 521	19	0.9 196832	70	AC027054
C 449	19	0.9 172000	68	AC024194	C 522	19	0.9 197577	80	AL358791
C 450	19	0.9 172187	81	AL390879	C 523	19	0.9 197735	83	CEY6647
C 451	19	0.9 172256	72	AC051637	C 524	19	0.9 197772	86	AC008129
C 452	19	0.9 172496	68	AC023968	C 525	19	0.9 197846	66	AC021126
C 453	19	0.9 172533	87	AC009116	C 526	19	0.9 197961	70	AC026890
C 454	19	0.9 172622	68	AC024065	C 527	19	0.9 198177	13	ANCHRTV17
C 455	19	0.9 172809	75	AC073834	C 528	19	0.9 198237	70	AC026299
C 456	19	0.9 172822	62	AC012429	C 529	19	0.9 198657	81	AL442648
C 457	19	0.9 172976	68	AC023843	C 530	19	0.9 199099	80	AL357974
C 458	19	0.9 172990	62	AC011973	C 531	19	0.9 199174	86	AC007742
C 459	19	0.9 173360	84	CNS07ECU	C 532	19	0.9 199414	77	AC084742
C 460	19	0.9 173564	4	AC007053	C 533	19	0.9 199577	13	ANCHRTV57
C 461	19	0.9 174450	79	AL354754	C 534	19	0.9 199749	13	ANCHRTV68
C 462	19	0.9 174453	69	AC025953	C 535	19	0.9 199987	13	ANCHRTV15
C 463	19	0.9 174570	62	AC012449	C 536	19	0.9 200333	74	AC073279
C 464	19	0.9 175302	76	AC079350	C 537	19	0.9 200987	75	AC078956
C 465	19	0.9 176168	61	AC010798	C 538	19	0.9 201179	75	AC079305
C 466	19	0.9 176169	66	AC021513	C 539	19	0.9 201302	77	AC087644
C 467	19	0.9 177036	78	AL139136	C 540	19	0.9 201302	77	AC087644
C 468	19	0.9 177163	92	HS134E15	C 541	19	0.9 202632	72	AC062029
C 469	19	0.9 177212	69	AC025733	C 542	19	0.9 203251	62	AC011635
C 470	19	0.9 177317	87	AC020637	C 543	19	0.9 203258	87	AC010482
C 471	19	0.9 177902	69	AC025936	C 544	19	0.9 206187	74	AC069139
C 472	19	0.9 178109	86	AC006079	C 545	19	0.9 206954	60	AC008367
C 473	19	0.9 178295	66	AC021798	C 546	19	0.9 207410	66	AC021130
C 474	19	0.9 178788	81	AL1391816	C 547	19	0.9 207722	65	AC019206
C 475	19	0.9 178825	82	AP001203	C 548	19	0.9 207944	64	AC016999
C 476	19	0.9 179007	62	AP0012275	C 549	19	0.9 207945	91	CNS010SS
C 477	19	0.9 179413	88	AC023486	C 550	19	0.9 209575	75	AC073739
C 478	19	0.9 179919	61	AL1390875	C 551	19	0.9 210000	60	AC004555
C 479	19	0.9 180017	74	AC073277	C 552	19	0.9 210824	65	AC019131
C 480	19	0.9 180073	74	AC073096	C 553	19	0.9 213847	80	AL359923
C 481	19	0.9 180240	68	AC023416	C 554	19	0.9 216300	75	AC069335
C 482	19	0.9 180512	81	AL450023	C 555	19	0.9 217141	75	AC073998
C 483	19	0.9 181261	81	AL1391730	C 556	19	0.9 217616	76	AC079531
C 484	19	0.9 181822	83	CNS01DV9	C 557	19	0.9 219801	62	AC011406
C 485	19	0.9 182062	80	AL356493	C 558	19	0.9 220967	62	AC012410
C 486	19	0.9 182328	82	AL1590307	C 559	19	0.9 232392	58	AP250284
C 487	19	0.9 182417	69	AC025291	C 560	19	0.9 232839	68	AP201647
C 488	19	0.9 182895	62	AC011078	C 561	19	0.9 234161	5	AE003769
C 489	19	0.9 183391	81	AL442645	C 562	19	0.9 237181	68	AC023460
C 490	19	0.9 184092	65	AC018715	C 563	19	0.9 237181	68	AC023460
C 491	19	0.9 184331	68	AC023651	C 564	19	0.9 244239	60	AC006880
C 492	19	0.9 185323	68	AC024399	C 565	19	0.9 254846	5	AE003725
C 493	19	0.9 185417	77	AC084356	C 566	19	0.9 254877	60	AC006904
C 494	19	0.9 186048	66	AC021621	C 567	19	0.9 254877	60	AC006904
C 495	19	0.9 186184	68	AC024646	C 568	19	0.9 257817	60	AC006909
C 496	19	0.9 186333	70	AC027355	C 569	19	0.9 261402	83	CEY41C4
C 497	19	0.9 186447	67	AC084368	C 570	19	0.9 264646	5	AE003818
C 498	19	0.9 186589	90	AL353734	C 571	19	0.9 268218	75	AC078989
C 499	19	0.9 186608	90	AL353734	C 572	19	0.9 268267	70	AC026731
C 500	19	0.9 188177	63	AC013400	C 573	19	0.9 268857	2	AP002560
C 501	19	0.9 188199	75	AC079242	C 574	19	0.9 269518	5	AE003751
C 502	19	0.9 188547	87	AC012066	C 575	19	0.9 276829	60	AC006741
C 503	19	0.9 189551	72	AC055118	C 576	19	0.9 276829	60	AC006741
C 504	19	0.9 190001	78	AL133541	C 577	19	0.9 279110	3	RPX03
C 505	19	0.9 190412	73	AC012485	C 578	19	0.9 282173	83	CEY40B1
C 506	19	0.9 190506	72	AC068990	C 579	19	0.9 291622	60	AC006796
C 507	19	0.9 190679	82	AL1590382	C 580	19	0.9 292390	60	AC006900
C 508	19	0.9 190793	71	AC034298	C 581	19	0.9 293219	76	AC079577
C 509	19	0.9 190882	89	AL133355	C 582	19	0.9 302797	4	AE003488
C 510	19	0.9 192318	75	AC074021	C 583	19	0.9 306131	60	AC006874
C 511	19	0.9 192554	88	AC024170	C 584	19	0.9 311043	4	AC009888
C 512	19	0.9 193625	82	AP001194	C 585	19	0.9 329362	5	AE003681
C 513	19	0.9 193758	69	AC023570	C 586	19	0.9 339485	13	AP172282
C 514	19	0.9 194695	71	AC044835	C 587	19	0.9 340000	92	HS21C003
C 515	19	0.9 195251	68	AC023862	C 588	18	0.9 32	9	AI14206
C 516	19	0.9 195251	68	AC023862	C 589	18	0.9 126	54	G43581
C 517	19	0.9 196317	64	AC016766	C 590	18	0.9 131	92	HS66H6R
C 518	19	0.9 196416	86	AC006483	C 591	18	0.9 142	93	HS288504
C 519	19	0.9 196430	74	AC073424	C 592	18	0.9 171	7	SHRPERC

593	18	0.9	207	91	AY013378	AY013378	Saguinus	666	18	0.9	4382	96	PFC04084	AL010254	Plasmodiu
594	18	0.9	208	91	AY013379	AY013379	Saguinus	667	18	0.9	4729	4	AB040460	AB040460	Wetarsenia
595	18	0.9	340	54	DM83E10T	Z50675	D. melanoga	668	18	0.9	4912	8	CER277745	AJ277745	Colurnix
596	18	0.9	406	54	H0M0T2438	L18097	Human STS U	669	18	0.9	4962	8	FR15172	Y15172	Fugu rubrip
597	18	0.9	448	54	G60780	G60780	SHGC-81965	670	18	0.9	5114	6	CERY47H9A	AL031636	Caenorhab
598	18	0.9	451	54	G60755	G60755	SHGC-81932	671	18	0.9	5356	85	AB037744	AB037744	Hom sapi
599	18	0.9	552	91	H0S223253	AJ223253	Hom sapi	672	18	0.9	5404	96	SGR297711	AJ297711	Sphaerich
600	18	0.9	609	54	DM164H2S	Z31909	Drosophila	673	18	0.9	6207	64	AC017165	AC017165	Drosophila
601	18	0.9	643	9	AB6363	AB6363	Sequence 10	674	18	0.9	6244	15	ST0011844	AJ011844	Solanum t
602	18	0.9	643	10	E66381	E66381	Genome DNA	675	18	0.9	6300	3	SSARP1	X59128	Synechocyst
603	18	0.9	660	14	CNS01CXY	AL115370	Bovine	676	18	0.9	6355	94	AF314088	AF314088	Mus muscu
604	18	0.9	746	7	AF087937	AF087937	Sus scrofa	677	18	0.9	6813	6	CERY75B7AR	AC024870	Caenorhab
605	18	0.9	817	53	CNS06ETT	AL395559	T3 end of	678	18	0.9	6866	97	HSU1566	Z70039	Human s
606	18	0.9	877	53	CNS06ONP	AL408299	T7 end of	679	18	0.9	7858	13	AF276703	AF276703	Oryza sat
607	18	0.9	882	94	AF017453	AF017453	Mus muscu	680	18	0.9	8460	6	CEY52D3	CERY52D3	Caenorhab
608	18	0.9	888	8	L0U17057	U17057	Lepistosteus	681	18	0.9	10085	12	AB015471	AB015471	Arabidops
609	18	0.9	900	74	AC054752	AC054752	Gardia i	682	18	0.9	10124	4	AC087080	AC087080	Caenorhab
610	18	0.9	937	72	AC061474	AC061474	Gardia i	683	18	0.9	10846	1	AE005144	AE005144	Halobacte
611	18	0.9	943	72	AC061474	AC061474	Gardia i	684	18	0.9	11042	3	MTV045	MTV045	Myobacte
612	18	0.9	946	72	AC061474	AC061474	Gardia i	685	18	0.9	11831	83	CERY43D4_4	AL022000	Halobacte
613	18	0.9	948	53	CNS06PLD	AL409511	T7 end of	686	18	0.9	12787	1	AE005170	AE005170	Halobacte
614	18	0.9	948	54	CNS07DXY	AL441068	T7 end of	687	18	0.9	15469	60	AC009067	AC009067	Hom sapi
615	18	0.9	957	53	CNS0608S	AL407766	T3 end of	688	18	0.9	16332	1	AF104912	AF104912	Escherich
616	18	0.9	1039	91	BC005211	BC005211	Hom sapi	689	18	0.9	16343	6	CER24D3	CER24D3	Caenorhab
617	18	0.9	1129	94	AB02526353	AB025265	Mus muscu	690	18	0.9	17137	6	DRONORCH03	K03508	D. melanoga
618	18	0.9	1130	5	AF033921	AF033921	Bacterioce	691	18	0.9	17411	3	MYCRRN1	L35043	Mycoplasma
619	18	0.9	1180	9	AX059937	AX059937	Sequence	692	18	0.9	17559	94	AB009692	AB009692	Mus sp. g
620	18	0.9	1206	14	KMAR52G	Z31563	K. merxianus	693	18	0.9	18245	65	AC017745	AC017745	Drosophila
621	18	0.9	1285	9	AX059885	AX059885	Sequence	694	18	0.9	19567	4	AC084157	AC084157	Caenorhab
622	18	0.9	1651	14	LEDI4RE	Z18277	Lycopersico	695	18	0.9	20919	8	FRU92572	FRU92572	Fugu rubrip
623	18	0.9	1763	91	BC005857	BC005857	Hom sapi	696	18	0.9	21520	4	AC024813	AC024813	Caenorhab
624	18	0.9	1767	88	AF035303	AF035303	Hom sapi	697	18	0.9	23712	12	AC079282	AC079282	Arabidops
625	18	0.9	1783	91	BC000951	BC000951	Hom sapi	698	18	0.9	24085	6	CERF56C4	CERF56C4	Caenorhab
626	18	0.9	1830	8	CC0271052	AJ271052	Coturnix	699	18	0.9	24123	1	AF068415	AF068415	Acidovora
627	18	0.9	1835	89	AK025146	AK025146	Hom sapi	700	18	0.9	24330	6	CERY49F6A	AC024800	Caenorhab
628	18	0.9	1851	13	AF266073	AF266073	Gomphus v	701	18	0.9	24709	5	CEC26H9A	Z36238	Caenorhab
629	18	0.9	1868	13	AF266074	AF266074	Gomphus e	702	18	0.9	24957	6	CER74	Z36238	Caenorhab
630	18	0.9	1872	13	AF266072	AF266072	At1gcmphu	703	18	0.9	26760	65	AC011857	AC011857	Drosophila
631	18	0.9	1910	2	BF091842	U91842	Bacillus fl	704	18	0.9	26767	6	CELM0789	AF040658	Caenorhab
632	18	0.9	1945	89	AK022700	AK022700	Hom sapi	705	18	0.9	26946	90	AL365328	AL365328	Hom sapi
633	18	0.9	1954	91	BC001253	BC001253	Hom sapi	706	18	0.9	27299	90	AL139810	AL139810	Human DNA
634	18	0.9	1960	88	AF102851	AF102851	Hom sapi	707	18	0.9	27493	6	CEIK10B4	AF025463	Caenorhab
635	18	0.9	1995	97	HS041806	U41806	Human BR13-	708	18	0.9	27874	63	AC014049	AC014049	Drosophila
636	18	0.9	1999	88	AP063604	AP063604	Hom sapi	709	18	0.9	29047	6	CER05A10	Z68108	Caenorhab
637	18	0.9	2021	91	BC003139	BC003139	Hom sapi	710	18	0.9	29523	4	AC024856	AC024856	Caenorhab
638	18	0.9	2035	12	AF040145	AF040145	Capanea g	711	18	0.9	30782	4	CERF3A8	Z81525	Caenorhab
639	18	0.9	2060	97	HS046751	U46751	Human phosp	712	18	0.9	32097	63	AC013173	AC013173	Drosophila
640	18	0.9	2083	9	AR077144	AR077144	Sequence	713	18	0.9	32209	6	CERJ24K24	AC006663	Caenorhab
641	18	0.9	2113	7	AB021664	AB021664	Sus scrofa	714	18	0.9	33700	6	CERF43D9	Z35640	Caenorhab
642	18	0.9	2136	89	AK000101	AK000101	Hom sapi	715	18	0.9	33038	2	AF222753	AF222753	Bradyniz
643	18	0.9	2253	5	AF032396	AF032396	Hyalophor	716	18	0.9	33455	6	CER04A8	Z82069	Caenorhab
644	18	0.9	2261	14	D88391	D88391	Oryza sativ	717	18	0.9	33516	6	CERF32A7	Z83107	Caenorhab
645	18	0.9	2279	8	BC001874	BC001874	Hom sapi	718	18	0.9	33539	6	CERF54E7	U00067	Caenorhab
646	18	0.9	2301	8	PAPPEBP	X72069	P. amurensi	719	18	0.9	33921	6	CER09D10	Z97185	Caenorhab
647	18	0.9	2317	14	CASAP6G	Z30193	C. albicans	720	18	0.9	34334	6	CERF49F1	AF100656	Caenorhab
648	18	0.9	2446	94	AB025269	AB025269	Mus muscu	721	18	0.9	34544	6	CERY58G8A	AC006808	Caenorhab
649	18	0.9	2466	94	AB025269	AB025269	Mus muscu	722	18	0.9	34598	6	CERL09D3	U64835	Caenorhab
650	18	0.9	2483	6	DDI13A	L08391	Dictyostell	723	18	0.9	34791	6	CERL04C9	U80955	Caenorhab
651	18	0.9	2503	89	AK025741	AK025741	Hom sapi	724	18	0.9	35079	6	CERF35F10	AF002198	Caenorhab
652	18	0.9	2512	93	HSMB02462	AL157427	Hom sapi	725	18	0.9	35094	90	AL158090	AL158090	Human DNA
653	18	0.9	2516	93	AF221892	AF221892	Hom sapi	726	18	0.9	35435	83	CER06F14	Z66524	Caenorhab
654	18	0.9	2550	2	CHITNCA	L35036	Chlamydomon	727	18	0.9	35807	6	CER113H5	U64842	Caenorhab
655	18	0.9	2678	89	AF221993	AF221993	Hom sapi	728	18	0.9	36298	6	CERF25B4	AC014943	Drosophila
656	18	0.9	2760	96	TERMIH8A	M61689	Tetrahymena	729	18	0.9	36569	63	AC014943	AC014943	Drosophila
657	18	0.9	2784	89	AK025288	AK025288	Hom sapi	730	18	0.9	36583	6	LMFL2464	AL365154	Leishmani
658	18	0.9	2806	13	AF266750	AF266750	Schistosac	731	18	0.9	36855	84	HS422E10	AL024496	Hom sapi
659	18	0.9	3440	12	AF005040	AF005040	Neurospor	732	18	0.9	36912	6	CERL05F6	AF040653	Caenorhab
660	18	0.9	3497	3	MAR300838	AJ300838	Metanospo	733	18	0.9	36926	63	AC015225	AC015225	Drosophila
661	18	0.9	4068	89	AH00730052	AP104015	Hom sapi	734	18	0.9	37010	5	CBRG28M15	AC084544	Caenorhab
662	18	0.9	4127	72	AC0057308	AC0057308	Hom sapi	735	18	0.9	37037	6	CER17C11	Z72507	Caenorhab
663	18	0.9	4247	12	AB049199	AB049199	Populus a	736	18	0.9	37142	69	AC025045	AC025045	Hom sapi
664	18	0.9	4301	93	HS4270775	AJ270775	Hom sapi	737	18	0.9	37545	6	CER21C9	Z73098	Caenorhab
665	18	0.9	4349	5	AF286897	AF286897	Plasmodiu	738	18	0.9	37680	80	AL359985	AL359985	Hom sapi

C 739	18	0.9	37854	5	CEC24H11	Z81475 Caenorhabd	C 812	18	0.9	74687	13	ATF1N13	AL391145 Arabidops
C 740	18	0.9	38514	5	CEC27A7	Z81041 Caenorhabd	813	18	0.9	74813	92	HS931K24	AL034430 Homo sapi
C 741	18	0.9	38627	6	CELC01G5	U50068 Caenorhabd	814	18	0.9	74836	64	AC016420	AC016440 Homo sapi
C 742	18	0.9	38856	6	CEY43F8A	AL032640 Caenorhab	C 815	18	0.9	75003	4	AC084444	AC084444 Caenorhab
C 743	18	0.9	39243	83	CEY81G3_4	Continuation (5 of	C 816	18	0.9	75192	61	AC009215	AC009225 Drosophill
C 744	18	0.9	39356	6	CEY32A11	Z81521 Caenorhabd	C 817	18	0.9	76114	90	AL445187	AL445187 Homo sapi
C 745	18	0.9	39873	83	CEM03A8	Z82066 Caenorhabd	C 818	18	0.9	76485	4	AC024872	AC024872 Caenorhab
C 746	18	0.9	40050	6	CELC23E4	L23651 C. elegans	C 819	18	0.9	76485	6	CELY76H12C	AC024211 Caenorhab
C 747	18	0.9	40053	6	CEY39G8C	AL032634 Caenorhab	C 820	18	0.9	77298	12	AB011483	AB011483 Arabidops
C 748	18	0.9	40108	97	HS065744	U65744 Human Chrom	C 821	18	0.9	77305	66	AC021228	AC021228 Homo sapi
C 749	18	0.9	40242	15	SPAC1F5	Z68136 S. pombe ch	C 822	18	0.9	76278	79	AL161741	AL161741 Homo sapi
C 750	18	0.9	40267	97	HSU15B9	Z69337 Human DNA s	C 823	18	0.9	78244	78	AL133337	AL133337 Homo sapi
C 751	18	0.9	40289	87	CEY48G10A	AC008991 Homo sapi	C 824	18	0.9	79826	4	AC002443	AC002443 Drosophill
C 752	18	0.9	40289	6	CEY48G10A	AL132847 Caenorhab	C 825	18	0.9	80552	90	AL356377	AL356377 Homo sapi
C 753	18	0.9	40352	3	SCG61A	AL135655 Streptomy	C 826	18	0.9	80837	93	HS0129L7	HS0129L7 Homo sapi
C 754	18	0.9	40468	6	CELC06B10	AF040654 Caenorhab	C 827	18	0.9	81970	71	AC041028	AC041028 Homo sapi
C 755	18	0.9	40641	4	AC006816	AC006816 Caenorhab	C 828	18	0.9	81970	71	AC041028	AC041028 Homo sapi
C 756	18	0.9	40662	6	CEY32B12B	AL031632 Caenorhab	C 829	18	0.9	82360	14	AP000606	AP000606 Arabidops
C 757	18	0.9	40780	6	CEM03C11	Z49128 Caenorhabd	C 830	18	0.9	82360	14	ATT31B5	ATT31B5 Arabidops
C 758	18	0.9	40933	6	DMC140G11	AL035395 Drosophill	C 831	18	0.9	82986	83	CENAX2	CENAX2 Arabidops
C 759	18	0.9	41369	6	CEM03C1	Z82279 Caenorhabd	C 832	18	0.9	83297	6	CELY71F9B	CELY71F9B Homo sapi
C 760	18	0.9	41381	5	CBRG46I19	AC084655 Caenorhab	C 833	18	0.9	83297	6	CELY71F9B	CELY71F9B Homo sapi
C 761	18	0.9	41552	5	CEC17D12	Z81473 Caenorhabd	C 834	18	0.9	83661	91	AP001439	AP001439 Homo sapi
C 762	18	0.9	41750	14	NCB7N4	AL390218 Neurospor	C 835	18	0.9	83661	91	AP001439	AP001439 Homo sapi
C 763	18	0.9	42282	6	CELC0304	U39472 Caenorhabd	C 836	18	0.9	84371	83	AP001963	AP001963 Arabidops
C 764	18	0.9	42446	12	AF010283	AF010283 Sorghum b	C 837	18	0.9	84371	83	AC005496	AC005496 Arabidops
C 765	18	0.9	42793	14	SC9168	Z38061 S. cerevisia	C 838	18	0.9	84601	12	ATT15C9	ATT15C9 Arabidops
C 766	18	0.9	44340	77	AC087105	AC087105 Dnlo rer	C 839	18	0.9	84785	13	AP002064	AP002064 Homo sapi
C 767	18	0.9	44508	65	AC020529	AC020529 Drosophill	C 840	18	0.9	84785	13	AP002064	AP002064 Homo sapi
C 768	18	0.9	45189	6	CELC34P11	U46755 Caenorhabd	C 841	18	0.9	84884	86	AC006130	AC006130 Homo sapi
C 769	18	0.9	45770	6	CELC05C11	AF125446 Caenorhab	C 842	18	0.9	85590	92	HS453D15	HS453D15 Homo sapi
C 770	18	0.9	46007	52	HS2A11	AL074873 Homo sapi	C 843	18	0.9	86324	13	ATAC010870	ATAC010870 Arabidops
C 771	18	0.9	47334	90	AL391843	AL035666 Human DNA	C 844	18	0.9	86324	13	AB005245	AB005245 Arabidops
C 772	18	0.9	47972	90	CELY31B43	AL391843 Human DNA	C 845	18	0.9	87459	12	AB015477	AB015477 Arabidops
C 773	18	0.9	48110	6	CELY32G9A	AC087794 Caenorhab	C 846	18	0.9	87805	83	CEY37B9_3	CEY37B9_3 Arabidops
C 774	18	0.9	48338	83	CEM02D9	Z81580 Caenorhabd	C 847	18	0.9	87805	83	CEY37B9_3	CEY37B9_3 Arabidops
C 775	18	0.9	49208	90	AL390298	AL390298 Human DNA	C 848	18	0.9	87878	85	AC007659	AC007659 Homo sapi
C 776	18	0.9	49217	65	AL390825	AL390825 Drosophill	C 849	18	0.9	88010	14	ATT14D3	ATT14D3 Arabidops
C 777	18	0.9	49623	4	AC087232	AC087232 Caenorhab	C 850	18	0.9	88401	14	P23M19	P23M19 Arabidops
C 778	18	0.9	50192	12	AC007195	AC007195 Arabidops	C 851	18	0.9	88682	6	CEY113G7B	CEY113G7B Arabidops
C 779	18	0.9	51757	86	AC007025	AC007025 Homo sapi	C 852	18	0.9	89245	12	AC006233	AC006233 Arabidops
C 780	18	0.9	51759	78	AC090787	AC090787 Homo sapi	C 853	18	0.9	89988	14	ATT22D6	ATT22D6 Arabidops
C 781	18	0.9	53371	76	AC084186	AC084186 Homo sapi	C 854	18	0.9	90130	15	TM021B04	TM021B04 Arabidops
C 782	18	0.9	54528	12	AB020754	AB020754 Arabidops	C 855	18	0.9	90130	15	TM021B04	TM021B04 Arabidops
C 783	18	0.9	54810	73	AC068732	AC068732 Homo sapi	C 856	18	0.9	90736	79	AL355868	AL355868 Homo sapi
C 784	18	0.9	55284	89	AL136318	AL136318 Human DNA	C 857	18	0.9	90736	79	AL355868	AL355868 Homo sapi
C 785	18	0.9	58753	6	CEY25C1A	AF125459 Caenorhab	C 858	18	0.9	92459	60	AC003118	AC003118 Homo sapi
C 786	18	0.9	59321	6	CEY43D4A	AL132846 Caenorhab	C 859	18	0.9	92459	60	AC003118	AC003118 Homo sapi
C 787	18	0.9	63527	74	AC069116	AC069116 Homo sapi	C 860	18	0.9	92745	84	PFMALAP1_3	PFMALAP1_3 Arabidops
C 788	18	0.9	63551	63	AC012978	AC012978 Drosophill	C 861	18	0.9	93290	12	AC0079279	AC0079279 Arabidops
C 789	18	0.9	64201	85	AC000054	AC000054 Homo sapi	C 862	18	0.9	93290	12	AC0079279	AC0079279 Arabidops
C 790	18	0.9	64508	72	AC060769	AC060769 Homo sapi	C 863	18	0.9	94257	61	AC010963	AC010963 Arabidops
C 791	18	0.9	64610	78	AC090753	AC090753 Homo sapi	C 864	18	0.9	94257	61	AC010963	AC010963 Arabidops
C 792	18	0.9	64610	78	AC090753	AC090753 Homo sapi	C 865	18	0.9	94257	61	AC010963	AC010963 Arabidops
C 793	18	0.9	65331	78	AC091337	AC091337 Homo sapi	C 866	18	0.9	95668	61	AC010240	AC010240 Homo sapi
C 794	18	0.9	65331	78	AC091337	AC091337 Homo sapi	C 867	18	0.9	95668	61	AC010240	AC010240 Homo sapi
C 795	18	0.9	65331	78	AC091337	AC091337 Homo sapi	C 868	18	0.9	95668	61	AC010240	AC010240 Homo sapi
C 796	18	0.9	67670	4	AE002910	AE002910 Drosophill	C 869	18	0.9	96785	62	ATT25P22	ATT25P22 Arabidops
C 797	18	0.9	68133	78	AC006575	AC006575 Homo sapi	C 870	18	0.9	96785	62	ATT25P22	ATT25P22 Arabidops
C 798	18	0.9	68763	4	AC0069718	AC0069718 Homo sapi	C 871	18	0.9	98029	85	AC004082	AC004082 Homo sapi
C 799	18	0.9	68948	14	ATT22N19	AL161357 Arabidops	C 872	18	0.9	98070	88	AC025463	AC025463 Homo sapi
C 800	18	0.9	69572	85	AC004769	AC004769 Homo sapi	C 873	18	0.9	98527	12	AC004044	AC004044 Arabidops
C 801	18	0.9	69709	6	CEY6B3B	AL032655 Caenorhab	C 874	18	0.9	98527	12	AC004044	AC004044 Arabidops
C 802	18	0.9	70592	83	AP002017	AP002017 Homo sapi	C 875	18	0.9	99231	12	AC005267	AC005267 Arabidops
C 803	18	0.9	70667	13	AC004680	AC004680 Arabidops	C 876	18	0.9	100000	91	AP000143	AP000143 Homo sapi
C 804	18	0.9	71327	13	AP001300	AP001300 Arabidops	C 877	18	0.9	100000	91	AP000143	AP000143 Homo sapi
C 805	18	0.9	71704	73	AC068468	AC068468 Homo sapi	C 878	18	0.9	100634	89	AP001594	AP001594 Homo sapi
C 806	18	0.9	71813	6	DMBNAL24	AL121805 Drosophill	C 879	18	0.9	101371	91	AC002534	AC002534 Arabidops
C 807	18	0.9	72719	78	AL354832_3	Continuation (4 of	C 880	18	0.9	101640	6	CELY4E10A	CELY4E10A Arabidops
C 808	18	0.9	72955	79	AL354832_3	Continuation (4 of	C 881	18	0.9	101640	6	CELY4E10A	CELY4E10A Arabidops
C 809	18	0.9	73000	78	AC090562	AC090562 Homo sapi	C 882	18	0.9	101840	73	AC068863	AC068863 Homo sapi
C 810	18	0.9	73031	92	HS78B3	Z82217 Human DNA s	C 883	18	0.9	102043	89	AF130351	AF130351 Homo sapi
C 811	18	0.9	73844	93	HSBK109D1	AL079330 Human DNA	C 884	18	0.9	102382	86	AC007557	AC007557 Homo sapi
									0.9	102507	93	HSJ324N14	HSJ324N14 Arabidops

C 885	18	0.9 103353	82	AP000672	AP000672 Homo sapi	C 958	18	0.9 117855	12	AC023279	AC023279 Genomic s
C 886	18	0.9 103587	81	AL512823	AL512823 Homo sapi	C 959	18	0.9 118334	85	AC005075	AC005075 Homo sapi
C 887	18	0.9 103750	82	AL590433	AL590433 Mus muscu	C 960	18	0.9 118425	90	AL356005	AL356005 Homo DNA
C 888	18	0.9 103970	64	AC015817-3	AC015817 Homo sapi	C 961	18	0.9 118636	82	AL590387	AL590387 Homo sapi
C 889	18	0.9 104148	86	AC006146	AC006146 Homo sapi	C 962	18	0.9 119393	85	AL590368	AL590368 Homo sapi
C 890	18	0.9 104834	73	AC068060	AC068060 Homo sapi	C 963	18	0.9 120593	13	AC003658	AC003658 Homo sapi
C 891	18	0.9 105080	88	AC026439	AC026439 Homo sapi	C 964	18	0.9 121151	92	AP002870	AP002870 Oryza sat
C 892	18	0.9 105733	12	AC012561	AC012561 Arabidops	C 965	18	0.9 121576	60	AC003117	AC003117 Homo sapi
C 893	18	0.9 105778	5	CE15	AL132943 Caenorhab	C 966	18	0.9 121597	91	AP000090	AP000090 Homo sapi
C 894	18	0.9 105815	13	AP002536	AP002536 Oryza sat	C 967	18	0.9 121990	68	AC003389	AC003389 Homo sapi
C 895	18	0.9 105906	62	AC011911	AC011911 Drosophila	C 968	18	0.9 122000	85	AC003093	AC003093 Homo sapi
C 896	18	0.9 106300	86	AC008602	AC008602 Homo sapi	C 969	18	0.9 122472	88	AC074264	AC074264 Homo BAC
C 897	18	0.9 106842	83	CEY11671-8	Continuation (9 of	C 970	18	0.9 122431	12	AC022314	AC022314 Homo sapi
C 898	18	0.9 106860	72	AC055703	AC055703 Mus muscu	C 971	18	0.9 122986	85	AC004915	AC004915 Homo sapi
C 899	18	0.9 107397	82	AP000612	AP000612 Homo sapi	C 972	18	0.9 123091	6	CEY8762A	CEY8762A
C 900	18	0.9 107931	12	AC005315	AC005315 Arabidops	C 973	18	0.9 123223	91	AB021816	AB021816 Homo sapi
C 901	18	0.9 108189	86	AC007128	AC007128 Homo sapi	C 974	18	0.9 123554	85	AB023049	AB023049 Homo sapi
C 902	18	0.9 109133	88	AC024571	AC024571 Homo sapi	C 975	18	0.9 123943	86	AC006208	AC006208 Homo sapi
C 903	18	0.9 109637	93	HS1842K24	AL050310 Human DNA	C 976	18	0.9 125041	73	AC068693	AC068693 Homo sapi
C 904	18	0.9 110000	74	AC073702-1	Continuation (2 of	C 977	18	0.9 125528	70	AC027389	AC027389 Homo sapi
C 905	18	0.9 110000	76	AC083893-3	Continuation (4 of	C 978	18	0.9 126356	60	AC006876	AC006876 Homo sapi
C 906	18	0.9 110000	77	AC087899-1	Continuation (2 of	C 979	18	0.9 126391	86	AC007243	AC007243 Homo sapi
C 907	18	0.9 110000	77	AC087899-2	Continuation (3 of	C 980	18	0.9 126800	80	AL365442	AL365442 Homo sapi
C 908	18	0.9 110000	79	AL35364-0	AL35364 Homo sapi	C 981	18	0.9 126865	88	AL362443	AL362443 Homo sapi
C 909	18	0.9 110000	80	AL359032-0	AL359032 Homo sapi	C 982	18	0.9 127037	87	AC011748	AC011748 Homo sapi
C 910	18	0.9 110000	83	CEY105C5-1	Continuation (3 of	C 983	18	0.9 127201	93	HS0604K5	HS0604K5
C 911	18	0.9 110000	83	CEY105C5-2	Continuation (3 of	C 984	18	0.9 127501	80	AL357852	AL357852 Homo sapi
C 912	18	0.9 110000	83	CEY113G7-2	Continuation (4 of	C 985	18	0.9 127544	66	AC020639	AC020639 Homo sapi
C 913	18	0.9 110000	83	CEY113G7-3	Continuation (6 of	C 986	18	0.9 127646	86	AC008171	AC008171 Homo sapi
C 914	18	0.9 110000	83	CEY116F11-5	Continuation (7 of	C 987	18	0.9 127652	82	AP001526	AP001526 Homo sapi
C 915	18	0.9 110000	83	CEY116F11-6	AL022596 Caenorhab	C 988	18	0.9 128230	92	HS13134	HS13134
C 916	18	0.9 110000	83	CEY37H9-0	Continuation (2 of	C 989	18	0.9 128388	85	AC003106	AC003106 Homo sapi
C 917	18	0.9 110000	83	CEY37H9-1	Continuation (3 of	C 990	18	0.9 128936	66	AC021265	AC021265 Homo sapi
C 918	18	0.9 110000	83	CEY39A1-1	Continuation (2 of	C 991	18	0.9 129171	90	AL445196	AL445196 Human DNA
C 919	18	0.9 110000	83	CEY39A1-2	Continuation (3 of	C 992	18	0.9 129254	90	AL353660	AL353660 Human DNA
C 920	18	0.9 110000	83	CEY39G8-0	Continuation (2 of	C 993	18	0.9 129654	92	HS71014	HS71014
C 921	18	0.9 110000	83	CEY43D4-1	Continuation (4 of	C 994	18	0.9 129849	92	AC078840	AC078840 Human DNA
C 922	18	0.9 110000	83	CEY43D4-2	Continuation (2 of	C 995	18	0.9 130059	12	AP003413	AP003413 Homo sapi
C 923	18	0.9 110000	83	CEY43D4-3	Continuation (2 of	C 996	18	0.9 130070	78	AP033234	AP033234 Homo sapi
C 924	18	0.9 110000	83	CEY47H9-0	Continuation (2 of	C 997	18	0.9 130674	83	AP003413	AP003413 Homo sapi
C 925	18	0.9 110000	83	CEY47H9-1	Continuation (3 of	C 998	18	0.9 131866	66	AC020834	AC020834 Mus muscu
C 926	18	0.9 110000	83	CEY47H9-2	Continuation (2 of	C 999	18	0.9 132200	79	AL161619	AL161619 Homo sapi
C 927	18	0.9 110000	83	CEY47H9-3	Continuation (2 of	C 1000	18	0.9 132200	79	AL161619	AL161619 Homo sapi
C 928	18	0.9 110000	83	CEY48G10-0	Continuation (2 of						
C 929	18	0.9 110000	83	CEY48G10-1	Continuation (2 of						
C 930	18	0.9 110000	83	CEY53F4-1	Continuation (2 of						
C 931	18	0.9 110000	83	CEY53F4-2	Continuation (3 of						
C 932	18	0.9 110000	83	CEY53F4-3	Continuation (4 of						
C 933	18	0.9 110000	83	CEY7A10-1	Continuation (3 of						
C 934	18	0.9 110000	83	CEY7A10-2	Continuation (3 of						
C 935	18	0.9 110000	83	CEY81G3-1	Continuation (2 of						
C 936	18	0.9 110000	83	CEY81G3-2	Continuation (4 of						
C 937	18	0.9 110000	84	PFMA13P2-2	Continuation (3 of						
C 938	18	0.9 111222	12	AC007764	AC007764 Genomic s						
C 939	18	0.9 111330	90	AL353810	AL353810 Human DNA						
C 940	18	0.9 111351	82	AP000725	AP000725 Homo sapi						
C 941	18	0.9 112022	92	HS4W9611	AP0009611 Homo sapi						
C 942	18	0.9 112339	92	HS110F11	AL030526 Human DNA						
C 943	18	0.9 112392	85	AC004028	AC004028 Human PAC						
C 944	18	0.9 112659	87	AC010677	AC010677 Homo sapi						
C 945	18	0.9 113547	62	AC011346	AC011346 Homo sapi						
C 946	18	0.9 113800	12	AC006567	AC006567 Arabidops						
C 947	18	0.9 114226	62	AC004710	AC004710 Homo sapi						
C 948	18	0.9 114339	86	AC006409	AC006409 Homo sapi						
C 949	18	0.9 115177	83	CEY12812	AC012812 Homo sapi						
C 950	18	0.9 115355	83	CEY55D9	AL008876 Caenorhab						
C 951	18	0.9 115762	82	AP000663	AP000663 Homo sapi						
C 952	18	0.9 116444	69	AC025440	AC025440 Homo sapi						
C 953	18	0.9 117006	62	AC001415	AC001415 Homo sapi						
C 954	18	0.9 117069	60	AC008907	AC008907 Homo sapi						
C 955	18	0.9 117071	78	AL133375	AL133375 Homo sapi						
C 956	18	0.9 117262	81	AC010314	AC010314 Homo sapi						
C 957	18	0.9 117367	87	AC010374	AC010374 Homo sapi						

## ALIGNMENTS

RESULT 1  
AF248966  
DEFINITION Homo sapiens HT028 mRNA, complete cds.  
ACCESSION AF248966  
VERSION AF248966.1 GI:12005668  
KEYWORDS FLI\_CDNA.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2049)  
Yang, Y., Xu, X., Gao, G., Xiao, H., Chen, Z. and Han, Z.  
Direct Submission  
Submitted (27-MAR-2000) Chinese National Human Genome Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, People's Republic of China  
LOCATION/Qualifiers  
1..2049  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/rname="HT028"  
/rstart=1  
/rstop=2049  
/product="HT028"



Db 1932 TCAACTTCATCTATACGATATATTAATGGAGTGGGAGATTCTACTTTTATGTTGSA 1991  
 1981 gtgagcaatgctcatcaagagtagacaataaagtaatgatlattccaasa 2032  
 Db 1992 GGGAGCAATGTCTATCAAGAGTGTACAAATTAAGTTATATATATTTCCAAAA 2043

RESULT	2				
LOCUS	HSMB00272				
DEFINITION	HSM800272	1884 bp	mRNA	PRI	18-FEB-2000
ACCESSION	Homo sapiens mRNA; CDNA	DKFpZ564O0582 (from clone DKFpZ564O0582);			
VERSION	partial cds.				
	AL049929				
	AL049929.1	GI:4884174			

polyA_signal	1854	.1859		
polyA_site	1874			
BASE COUNT	555 a	322 c	393 g	614 t
ORIGIN				

Query Match	87.48;	Score 1776;	DB 93;	Length 1884;
Best Local Similarity	99.98;	Pred. No. 0;		
Matches 1876; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0

Qy 155 gataatataaacaccaggctgttcttccgaatgaaattgagcctatracaggag 21

Db 1 gatatatgaataatcacaccaggctgttcttccgaatgaaattgagcctatracaggag 60

OY	215	agggatccaggaatgagctgattgttccaaatggagctctctgtgaaagaagcccttct	274
Db	61	AGCGATCCAGAGAGCGCGTCATTTGCCATGGCCTCTCTGTGAAAGAACCTTCTT	120
OY	275	ggccagagctcgcagttggaacctgttcatcgtctcctgggtaccgcgtcatgtgtg	334
Db	121	GGCGAGAGCTCGCAGTGGGTAACTGGTTTATCGTCCTCGGGCCACCGTCATGTGATG	180
OY	335	tgaaggagtgaaacaacacagctctaccaccaagcagtgtaattcgtacccttggaga	394
Db	181	TGAAGGAGTGAACAACATGGGCTACCCCGAGGCAAGTGTATTGTGTAACCTTTGAGA	240
OY	395	atgcagttccctttagtcttgaacagttgtgaaattcattcaactccatttctgaag	454
Db	241	ATCGAGTTCCTTTATGCTCTTGACAGTGTGCAAAATTTCAATTCATCTCTTTATTTCTAGG	300
OY	455	aaactcgtgtgttcttgcagtttgcagttccccaagtgaagaaagagtgatagtgaagaa	514
Db	301	AAACCTCGTGTGTTTGGCAGTTGGCTCCCGTAGAGAAAGGTATATGTAGAGAGAG	360
OY	515	caaatcagtgatttgaagaaccttcaatcaacttgcgcagctccgtatgcctgtttc	574
Db	361	CAAACTCAGTGTGGAAGACCTTCACATCACTTGGCCAGCTCCGTAAATGCCCTGTTT	420
OY	575	aagaaacctgttctcagttccactcctccccaattctctgaagtaggaacaatgaattg	634
Db	421	AGGAAATCTGTGTCATCTTCATCTCCCTCCCAATTTCTGTAGACGAACATGAACTTG	480
OY	635	accgtctcttcttcttgaactgaagtgatgcatacgaatttcaactgtcgtctgtc	694
Db	481	ACCTCTCTTTCTTTGTGAATGCAATGCTGTCAATGATATTTCAACTTGGCTGTGTT	540
OY	695	ataagacatcagcaagaagatattctccgtattatattcaacgaagctggcagtttg	754
Db	541	ATAAGCATCTAGCCAAGATCATTTCTCGATTATTAATCACTGGAGCTGGAGGTTTGG	600
OY	755	atgaattgggaagcgtatcgtgggaagactctgaacaatcagaatgtgtcttcaagatcc	814
Db	601	ATGAATTTGGGAAGCGTTATGTGGGAAACCTCTGAACAATTCAGATGTCTTCAAGATCC	660
OY	815	tgtgtgacgtctgcgaagaagcttgcagatgaacatctcaatcttctgtgtggaatgcag	874
Db	661	TGTGTGACGCTCTCGAAAAGTTTGCAAGTGTGACATGTAACTTTATGGTGGAAATCGAG	720
OY	875	tgttagagtttagtcaactgtgaagcaatttgaacctccctccattggaagcaagaagacta	934
Db	721	TGTAGATTAGTCACTGTGCAGATCTTTGAACTCCCTCATTTAGGAAGCAAGACTTA	780
OY	935	tccttgagcaaaaacagcgaagaccgcagcaagtcctcataacttgcataatgata	994
Db	781	TCCTTTGAGGCAAAACAGCAGAGAACCCAGCAAGCCCTTAACCTGTCATATAGATATA	840
OY	995	attcttgaatctcgtgtgtttcaacatgtaactttgataatgaatgcgcttggccttg	1054
Db	841	ATTTTGAATTTCCGTGTTTTCACATATGTAATCTTTGGATAATGATGCCCTTGGCCTTGG	900
OY	1055	ctgtgatattcacctctacaataatttgaacatgtaactcctgataatgaatcaatt	1114
Db	901	CTGTGATTATCACTCTTTACAAATTTTGGAAACATGCAATCTCGATATATATGCAATTT	960
OY	1115	atagatgaacaaacagagaattcogaatgtgaatgaatgaatccttgcagaatitgaaga	1174
Db	961	ATAGAGTGAACAAACAGAGATTCGAATGATGTAATGTTACTCGGCCAATATTTAGAA	1020
OY	1175	agggaggtcttgaagaatgagctgttttgtttaaataatacttttgttgcctttaaagaat	1234
Db	1021	AGGGGAGTGGAAATTTGGCTGTTTGTGTTAAATATATCTTTATGTGCTTTAAAGTAGAT	1080
OY	1235	agttacttcaacttalaaaaaaaatcaaatittgttcttatttgtgtgtgctgt	1294
Db	1081	AGTATCTTACATTTATATAAAAAAATCAAAATTTGTTCTTTATTTTGTGTGCGCTGT	1140



Oy	1295	gattcttctcgaagaaatattatagattgagocggaatcccaactcgtgattatagattccat	1354
Oy	1296		1355
Db	1141	gatgtgttttctcagagctgaattatagattgagacgtggaatcccatgctgattagattccat	1200
Oy	1355	aatatgcttgaatatactatgatataagacatttaataacatgattcattcctgtttaaaga	1414
Db	1201	aaatattgcttgtaataatttgaatgaatagaccattttaaatmacattgatttttcatttcgttttaatgc	1260
Oy	1415	atttggaaaatactgaacactgaagaagaactgtaaacaacttagaataagctcgttcttctggaaa	1474
Db	1261	atttggaaaattatgcactgaaggaaaatcttmaaacatttttggaaatagctcgtttatggaaaa	1320
Oy	1475	aagtgacactgaatttattatagaacaacttaagactgactgcttaactctttacacagataagt	1534
Db	1321	aaatgcactgaataattttatgacaaactctgcgaatgcgaatgcttttctttacacagatagct	1380
Oy	1535	gaaaatacatcttgggcctatgtaacttaactgaacaactcttgaagcttcaattgattgt	1594
Db	1381	gaataatcatgttttggcctattgtatatactatgacaaatttggtaaatgctttatatttgatgt	1440
Oy	1595	aaataactctgaacaacaaganaaaggcttttaacttaagtagtagccctaaataatgatagt	1654
Db	1441	aaatattctctgaaacaaagaaaaggcttttttaacttaagtagtagccctaaatgatagt	1500
Oy	1655	gcttaataatgccttagtcttcttggaaactgatactgaagtaacaagagacagctgttttcta	1714
Db	1501	gcttatattatattcccttagtcttcttggaaactgatactgaagtaacaagagacagctgttttcta	1560
Oy	1715	acctctctctgcaagtttgtctgtaacctatagctgagctataatgataactaaataactaat	1774
Db	1561	acctctctctctgcaagtttgtctgtaacctatagctgagctataatgataactaaataactaat	1620
Oy	1775	tgacttaagaagaagaactagacctcttggagtatataagatgacttcttcaatacacaaaa	1834
Db	1621	tgacttaagaagaagaagaactagacctcttggagtatataagatgacttcttcaatacacaaaa	1680
Oy	1835	atcccctgaaggacattttgagagcactgataataaaacatttttacttcaagtaacttccc	1894
Db	1681	atcccctgaaggacattttgagagcactgataataaaacatttttacttcaagtaacttccc	1740
Oy	1895	cctgtgtgaatttaactatggtcttctggtgataacttcaatcatalaagaatactaaagtgaagt	1954
Db	1741	cctgtgtgaatttaactatggtcttctggtgataacttcaatcatalaagaatactaaagtgaagt	1800
Oy	1955	gggtgtaatttactacttttatgtcttggagtgagaccaatgtctataaagatggacaataaag	2014
Db	1801	gggtgtaatttactacttttatgtcttggagtgagaccaatgtctataaagatggacaataaag	1860
Oy	2015	ttaatgatgatattcctcaaa	2032
Db	1861	ttaatgatgatattcctcaaa	1878
RESULT 3			
LOCUS	AC026156		
DEFINITION	Homo sapiens chromosome 3p clone RP11-169d16, WORKING DRAFT	HTG	28-MAR-2000
SEQUENCE	10 unordered pieces.		
ACCESSION	AC026156		
VERSION	AC026156.2 GI:7331308		
KEYWORDS	HTG: HTGS_PHASE1; HTGS_DNAFT.		
SOURCE	human.		
ORGANISM	Homo sapiens		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1 (bases 1 to 126543)		
AUTHORS	Lu,G., Zhao,Y., Lin,W., Dong,H., Wan,M., Xu,S., Gu,W., Tu,Y., Jia,J., Wu,C., Zhang,C., Zhong,M., Zhou,Y., Ren,S., Fu,G., Chen,Z and Huang M.		
TITLE	Direct Submission		
JOURNAL	Submitted (21-MAR-2000) Genomic Dept., Chinese National Human Genome Center at Shanghai, 315 Guo Shou Jing Road, Shanghai, Shanghai 201203, P. R. China		

	<p>On Mar 28, 2000 this sequence version replaced gi:721973.</p>					
	<p>-----Genome Center Information-----</p>					
	<p>Center: Chinese National Human Genome Center at Shanghai</p>					
	<p>Center Code: CHGC</p>					
	<p>Web Site: http://www.chgc.sh.cn</p>					
	<p>Email: mhuang@chgc.sh.cn or fujiang@chgc.sh.cn</p>					
	<p>-----End Genome Center Information-----</p>					
	<p>* NOTE: This is a "working draft" sequence. It currently consists of 10 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the configs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.</p>					
	<p>I            3411: contig of 3411 bp in length</p>					
	*	gap of unknown length				
	*	3412	7987: contig of 4576 bp in length			
	*	gap of unknown length				
	*	7988	16490: contig of 8503 bp in length			
	*	gap of unknown length				
	*	16491	30218: contig of 13728 bp in length			
	*	gap of unknown length				
	*	30219	44895: contig of 14677 bp in length			
	*	gap of unknown length				
	*	44896	55628: contig of 10733 bp in length			
	*	gap of unknown length				
	*	55629	71833: contig of 16205 bp in length			
	*	gap of unknown length				
	*	71834	84537: contig of 12704 bp in length			
	*	gap of unknown length				
	*	84538	103706: contig of 19189 bp in length			
	*	gap of unknown length				
	*	103707	126543: contig of 22837 bp in length.			
	<p>Location/Qualifiers</p>					
	<p>source</p>					
	<p>1..126543</p>					
	<p>/organism="Homo sapiens"</p>					
	<p>/db_xref="taxon:9606"</p>					
	<p>/chromosome="3p"</p>					
	<p>/clone="RP11-169o16"</p>					
	<p>BASE COUNT        34774 a   28466 c   28425 g   34834 t        44 others</p>					
	<p>ORIGIN</p>					
	<p>Query Match                      53.0%: Score 1078; DB 69; Length 126543;</p>					
	<p>Best Local Similarity     100.0%: Pred. No. 0;</p>					
	<p>Matches 1078; Conservative    0; Mismatches    0; Indels    0; Gaps    0;</p>					
Oy	954	gaagaaccgcaaglcacctaaacctcgcataaagatlaaatcttgaataatccgtgct	1013			
Db	81178	GAGAGACCACGACGACGCCCTATACCTTCGATTAAGTAAATTGGAAATAATCCGTGCT	81237			
Oy	1014	tttaacaatgatccttgataaagatgcgccttgccctggcctcgatcatcacctcta	1073			
Ddb	81238	TTTCAACATGCTACTTTTGGAATAAAGATGCCCTTGCGCTTGCGATTAACAACCTTAA	81297			
Oy	1074	caatatcttgacacatgacatcctcgcataatgataatgaatattatagatgacaaacagaa	1133			
Dd	81298	CATATTTTGGACATCATGATCCGATATGATAGCATATTTTATAGAGAACCAACCCAGA	81357			
Oy	1134	gattcgaatgatcgaatccttacctcgcacaaatctgaagaaggagtctggaatcgct	1193			
Dd	81358	GATTCGATGATGATTTGAATGATTACTGTCGCAAAATTAGAAAAGGGGATTTGGCT	81417			
Oy	1194	gttttgttaaaataatcctttagtctgtgctttaagfatagtatgaacttaacttaata	1253			
Dd	81418	GTTTTGTTAAAAAATATCTTTTAACTGTGCTTTTAAAGTAGATGATTAATCTTTACATTTATA	81477			
Oy	1254	aaaaaaaatcaatcttccttctatcttcgtctgctgcttgatctttctctagaatgaa	1313			
Dd	81478	AAAAAAAAATCAATTTTGTCTTTATTTGTGTGTCGCTGTGATGTTTCTAGAGTGAA	81537			
Oy	1314	tctatagatctgaacgtaaacctcactcgtggtatagatctccaataatagcttgaaatataag	1373			

Db	81538	TTATAGTTTACGTCGAATCCACCTGGTATAGATTCACAAATATATGCTGATATATATG	81597
Qy	1374	atatgacattatgaacattgatttattccttgcttatgaatttggaaatagacga	1433
Db	81598	ATTTACCATTTTATTAACATTGATTTTCTGTTTAATGCAATTTGGAAATATGCACTGA	81657
Qy	1434	aagaagatgaacatttgaatagctcgctgatgtaaaaaagtcacbaattatga	1493
Db	81658	AGAATAATTAACCATTTTGGATAGCTCGCTGTATGAAAAAGTGCATGATTTATTA	81717
Qy	1494	gaaacactagaatgcttcaactctcttcaacagatagaatgaatataatcttggcga	1553
Db	81718	GCAACAATCTTGAATCTTAACTTCTTTACACAGCATAGTGAATATATTTTGGGCTA	81777
Qy	1554	ttgtactatgacaatttggtaaatgcttcaatttgaatgaataactcgtgaacaga	1613
Db	81778	TTGTATCATATGAAACATTTTGAATGCTCTTAAATTTGATTAATTAACCTGAAACAGA	81837
Qy	1614	gaaaagttttaactagtagccctaaataatlgatgtagtattatataatgcctagt	1673
Db	81838	GAAGAAGTTTTTAATCTTGAAGTACCCCTTAATATGATGTGCTTAATATATGCTTAAT	81897
Qy	1674	tttgaactatcctcgtgagtaacagaggaacagctgctttttaaaccctcttcgtaattg	1733
Db	81898	TTTGGAACTGTATCTACGATACAGACAGACCTGTTTTTAACCTCTTCGCAAGTTTG	81957
Qy	1734	ttgacactacatggtgctaatatagataactaaataactacattgacttaagaagaactag	1793
Db	81958	TTGACCTTACATGGCGTAAATATGATACATTAATAATATACATTGATCTTAGAACAACCTG	82017
Qy	1794	cctgttgagatataatagatgcttttcaattatacacacaaaacccctgagggacaatttg	1853
Db	82018	CCCTTGGAGATATATGATAGTCTTTTATTATATACACAAAAATCCCTAGGGACATTTTG	82077
Qy	1854	aggcaatgaataataaacatttcaatttcaactgaactctcccccctgtaagttacatg	1913
Db	82078	AGCGATGATATATTAACATTTTATTTCAGTAACTTTTCCCTGTGTAACTTACATGG	82137
Qy	1914	tttgggttaacaacttcatcttcataagaataatgaatggaatggatggaacttcaatttta	1973
Db	82138	TTTGGGTGAACAATCTTCAATTCATATGAAATTAATGTGAACTGGGAATTCCTACTTTTGA	82197
Qy	1974	tgttggatggaacaatgtctcatcaagaatgacaataaagttatgtatctcaaa	2031
Db	82198	TGTTGAGATGGACCAATGTCTATCAAGAGACTGCACAATTAAGTTATATGATATTCGAA	82255
RESULT 4			
LOCUS	HSY17975	622 bp	mRNA
DEFINITION	Homo sapiens mRNA for putative vacuolar protein ATPase membrane sector associated protein M8-9.		
ACCESSION	Y17975		
VERSION	Y17975.1		
KEYWORDS	vacuolar proton ATPase membrane sector associated protein M8-9, human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
REFERENCE	1 (bases 1 to 622)		
AUTHORS	Ludwig, J., Kerscher, S., Brandt, U., Pfeiffer, K., Gellawi, F., Apps, D. K., and Schagger, H.		
TITLE	Identification and characterization of a novel 9.2-kDa membrane sector-associated protein of vacuolar proton-ATPase from chromaffin granules		
JOURNAL	J. Biol. Chem.	273 (18),	10939-10947 (1998)
MEDLINE	9625166		
REFERENCE	2 (bases 1 to 622)		
AUTHORS	Ludwig, J.H.		
TITLE	Direct Submission		
JOURNAL	Submitted (20- AUG-1998)	J.H. Ludwig, Universitätsklinikum Frankfurt, Gustav Embden-Zentrum der Biologischen Chemie (ZBC),	

COMMENT	FEATURES
Institut fuer Biochemie I, Theodor Stern-Kai 7 Haus 25 B, 60590 Frankfurt, FRG	Related entry: R78826.
	Location/Qualifiers
	1..622
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
CDS	192..494
	/codon_start=1
	/product="putative vacuolar proton ATPase membrane sector associated protein M8-9"
	/protein_id="CAI76984.1"
	/db_xref="GI:3451261"
	/db_xref="sp P88RBL:075787"
	/translation="MTSLTSGNNAVEELTVASGPTSLIKRTTLLEKQAKPPASPTINLAKINIEISVFNFMALMALALAVLITITSYINIMMDPGDITLYRMTNOKIRMD"
BASE COUNT	186 a 103 c 134 g 199 t
ORIGIN	

[illegible]

ACCESSION G27225  
 VERSION GI:1375475  
 KEYWORDS STS: STS sequence; primer; sequence tagged site.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 513)  
 AUTHORS Myers, R.M.  
 JOURNAL Unpublished (1995)  
 COMMENT  
 Contact: Richard M. Myers  
 Stanford Human Genome Center (SHGC)  
 Stanford University School of Medicine  
 Department of Genetics, M-344, Stanford, CA 94305, USA  
 Tel: 4157259687  
 Fax: 4157259689  
 Email: myers@shgc.stanford.edu  
 Primer A: CTTGATGACATTCGTCACATCC  
 Primer B: TTTTATTTCACTTTCCTCC  
 STS size: 128  
 PCR Profile:  
 Initial incubation: 94 degrees C for 90 seconds  
 Denaturation: 94 degrees C for 15 seconds  
 Annealing: 62 degrees C for 23 seconds  
 Polymerization: 72 degrees C for 30 seconds  
 PCR Cycles: 30  
 Thermal Cycler: Perkin Elmer 9600  
 Protocol:  
 Template: 25 ng  
 Primer: each 1 uM  
 dNTPs: each 200 uM  
 Tag Polymerase: 0.05 units/uM  
 Total Vol: 10 uL  
 Buffer:  
 MgCl2: 2.5 mM  
 KCl: 50 mM  
 Tris-HCl: 20 mM  
 pH: 8.3  
 Prepared with primer pairs provided by Sandoz, derived from R51314  
 -- Washington University/Merck EST sequence.  
 FEATURES  
 Location/Qualifiers  
 1..513  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="x"  
 STS  
 primer\_bind 38..165  
 primer\_bind complement(142..165)  
 BASE COUNT 150 a 97 c 84 g 177 t 5 others  
 ORIGIN  
 Query Match 12.5% Score 254 DB 54 Length 513;  
 Best Local Similarity 100.0% Prid. No. 1.8e-134 Indels 0 Gaps 0;  
 Matches 254 Conservative 0 Mismatches 0 Indels 0 Gaps 0;  
 Oy 1779 ctaagaagaactagcctgtgagatataagcttcattacatacacaaatcc 1838  
 |||||||  
 Db 259 CTAAAGAAAGACAGCCCTGTGAGATATAGATGCTTTCTATACACAAATTC 200  
 Oy 1839 ctgaagacatttggagcatataataaacttttatttcgtaactttccctcg 1898  
 |||||||  
 Db 199 CTGAGGAGCATTTTGAGGAGCATTAATTAATTTTATTTTCACTTTTCCCTG 140  
 Oy 1899 tctaatgactatgcttggtagtaacattcatatagaataatgaagtgagtg 1958  
 |||||||  
 Db 139 TGTAACTTACTATGTTGTGTACTCACTTCATTTCTATGATATTAAGTGAAGTGGT 80

Oy 1959 gaattctacttlatgttgagtgagcccaatgcatcaagagtgacaaataagttaa 2018  
 |||||||  
 Db 79 GAATCTACTTTTATGTTGAGTGGAGCAGCAATGCTTCTCAAGAGTGCACAAATAAGTTAA 20  
 Oy 2019 tgaatgttccaaa 2032  
 |||||||  
 Db 19 TGATGATTCACAAA 6  
 RESULT 6  
 AC026156/c  
 LOCUS Homo sapiens chromosome 3p clone RP11-169016, WORKING DRAFT  
 DEFINITION  
 AC026156.2 GI:7331308  
 ACCESSION  
 VERSION HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 126543)  
 AUTHORS Lu, G., Zhao, Y., Lin, W., Dong, H., Wan, M., Xu, S., Gu, W., Tu, Y.,  
 Jia, J., Wu, C., Zhang, C., Zhong, M., Zhou, Y., Ren, S., Fu, G., Chen, Z.  
 and Huang, M.  
 Direct Submission  
 Submitted (21-MAR-2000) Genomic Dept., Chinese National Human  
 Genome Center at Shanghai, 315 Guo Shou Jing Road, Shanghai,  
 Shanghai 201203, P. R. China  
 On Mar 28, 2000 this sequence version replaced gi:7271973.  
 -----Genome Center Information-----  
 Center: Chinese National Human Genome Center at Shanghai  
 Center Code: CHGC  
 Web site: http://www.chgc.sh.cn  
 Email: mhuang@chgc.sh.cn or fujang@chgc.sh.cn  
 -----End Genome Center Information-----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 10 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1 3411: contig of 3411 bp in length  
 \* 3412 7987: contig of 4576 bp in length  
 \* 7988 16490: contig of 8503 bp in length  
 \* 16491 30218: contig of 13728 bp in length  
 \* 30219 44895: contig of 14677 bp in length  
 \* 44896 55628: contig of 10733 bp in length  
 \* 55629 71833: contig of 16205 bp in length  
 \* 71834 84537: contig of 12704 bp in length  
 \* 84538 103706: contig of 19169 bp in length  
 \* 103707 126543: contig of 22837 bp in length.  
 Location/Qualifiers  
 1..126543  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="3p"  
 /clone="RP11-169016"  
 BASE COUNT 34774 a 28466 c 28425 g 34834 t 44 others  
 ORIGIN

Query Match 7.4%; Score 151; DB 69; Length 126543;  
Best Local Similarity 100.0%; Pred. No. 8e-75;  
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 684 gctgctgcataaagcattagcgaagatcctcctgattatatactgagct 743  
|||||  
Db 89809 GCTGCTGCTATACCATCTAGCCAGATCATCTCTGATTATATTCACGTGAGCT 89750  
|||||

Qy 744 ggcaggttgagatgaattgggaagcgttatggggaagcctcgaacattcgaagatgc 803  
|||||  
Db 89749 GGCAGCTTGGATGAATTGGGAACGCTTATGGGAAGACTGTGAACATTCAGAGATGC 89690  
|||||

Qy 804 ttctaagctcttgctgacgcctcgaagaag 834  
|||||  
Db 89689 TTCTAAGATCCTCTTGTGACGCTCTGCAAAAG 89659  
|||||

RESULT 7  
A74462 148 bp DNA PAT 15-OCT-1999  
LOCUS A74462 Sequence 148 from Patent WO9401548.  
DEFINITION A74462  
ACCESSION A74462  
VERSION A74462.1 GI:6064476  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 148)  
AUTHORS SIBSON,D.R. and GROSS,J.  
TITLE HUMAN NUCLEIC ACID FRAGMENTS, ISOLATED FROM BRAIN ADRENAL TISSUE,  
PLACENTA OR BONE MARROW  
PATENT: WO 9401548-A 148 20-JAN-1994;  
JOURNAL MEDICAL RES COUNCIL (GB); SIBSON DAVID ROSS (GB)  
FEATURES  
source  
1. 148  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 47 a 14 c 25 g 57 t 5 others  
ORIGIN

Query Match 5.7%; Score 116; DB 9; Length 148;  
Best Local Similarity 100.0%; Pred. No. 6.2e-55;  
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1145 attgaatgttaccctgtgcagaattagaagaagggtgtggaattgctgtttgttaa 1204  
|||||  
Db 14 ATTGAATGTTACCTGTGCCAGATTGAAAGGGGCTTGGAAATGGCTTTGTAA 73  
|||||

Qy 1205 atatacttttagtgctttaagtagatagtagtacttaccattataaaaaa 1260  
|||||  
Db 74 ATATATCTTTAGTGCTTTAAAGTAGATGATACCTTACATTTATATAAAAA 129  
|||||

RESULT 8  
A77441 148 bp DNA PAT 19-OCT-1999  
LOCUS A77441 Sequence 148 from Patent EP0587279.  
DEFINITION A77441  
ACCESSION A77441.1 GI:6089106  
VERSION A77441.1  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 148)  
AUTHORS SIBSON,D.R. and Hadfield,K.M.  
TITLE HUMAN NUCLEIC ACID FRAGMENTS ISOLATED FROM BRAIN, ADRENAL TISSUE,  
PLACENTA OR BONE MARROW AND THEIR USE  
PATENT: EP 0587279-A 148 16-MAR-1994;  
JOURNAL MEDICAL RES COUNCIL (GB)  
FEATURES  
source  
1. 148  
/organism="unidentified"

BASE COUNT 47 a 14 c 25 g 57 t 5 others  
ORIGIN

Query Match 5.7%; Score 116; DB 9; Length 148;  
Best Local Similarity 100.0%; Pred. No. 6.2e-55;  
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1145 attaaatgttaccctgtgcagaattagaagaagggtgtggaattgctgtttgttaa 1204  
|||||  
Db 14 ATTGAATGTTACCTGTGCCAGATTGAAAGGGGCTTGGAAATGGCTTTGTAA 73  
|||||

Qy 1205 atatacttttagtgctttaagtagatagtagtacttaccattataaaaaa 1260  
|||||  
Db 74 ATATATCTTTAGTGCTTTAAAGTAGATGATACCTTACATTTATATAAAAA 129  
|||||

RESULT 9  
A74460 151 bp DNA PAT 15-OCT-1999  
LOCUS A74460 Sequence 146 from Patent WO9401548.  
DEFINITION A74460  
ACCESSION A74460  
VERSION A74460.1 GI:6064474  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 151)  
AUTHORS SIBSON,D.R. and GROSS,J.  
TITLE HUMAN NUCLEIC ACID FRAGMENTS, ISOLATED FROM BRAIN ADRENAL TISSUE,  
PLACENTA OR BONE MARROW  
PATENT: WO 9401548-A 146 20-JAN-1994;  
JOURNAL MEDICAL RES COUNCIL (GB); SIBSON DAVID ROSS (GB)  
FEATURES  
source  
1. 151  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 49 a 14 c 25 g 58 t 5 others  
ORIGIN

Query Match 5.7%; Score 116; DB 9; Length 151;  
Best Local Similarity 100.0%; Pred. No. 6.2e-55;  
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1145 attgaatgttaccctgtgcagaattagaagaagggtgtggaattgctgtttgttaa 1204  
|||||  
Db 16 ATTGAATGTTACCTGTGCCAGATTGAAAGGGGCTTGGAAATGGCTTTGTAA 75  
|||||

Qy 1205 atatacttttagtgctttaagtagatagtagtacttaccattataaaaaa 1260  
|||||  
Db 76 ATATATCTTTAGTGCTTTAAAGTAGATGATACCTTACATTTATATAAAAA 131  
|||||

RESULT 10  
A77439 151 bp DNA PAT 19-OCT-1999  
LOCUS A77439 Sequence 146 from Patent EP0587279.  
DEFINITION A77439  
ACCESSION A77439  
VERSION A77439.1 GI:6089104  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 151)  
AUTHORS SIBSON,D.R. and Hadfield,K.M.  
TITLE HUMAN NUCLEIC ACID FRAGMENTS ISOLATED FROM BRAIN, ADRENAL TISSUE,  
PLACENTA OR BONE MARROW AND THEIR USE  
PATENT: EP 0587279-A 146 16-MAR-1994;  
JOURNAL MEDICAL RES COUNCIL (GB)  
FEATURES  
source  
1. 151  
/organism="unidentified"

BASE COUNT 49 a 14 c 25 g 58 t 5 others  
 ORIGIN

Query Match 5.7%; Score 116; DB 9; Length 151;  
 Best Local Similarity 100.0%; Pred. No. 6.2e-55;  
 Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1145 attgaatgcttgcgcgaattagaagaagg99gttgcgaattgctgttcttga 1204  
 |||||||  
 DB 16 ATTGAATGCTTACCTGGCCAGATTGAAAGGGGCTTGGAATGGCTGTTGTAAA 75  
 QY 1205 atatacttttgctgtcttaagtagatagatattacttataataaaaaa 1260  
 |||||||  
 DB 76 ATATATCTTTTACTGTGCTTTAAAGTAGATATACCTTACATTTATATAAAAAA 131

## RESULT 11

LOCUS A74463 204 bp DNA PAT 15-OCT-1999  
 DEFINITION Sequence 149 from Patent WO9401548.  
 ACCESSION A74463  
 VERSION A74463.1 GI:6064477  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.

REFERENCE 1 (bases 1 to 204)  
 AUTHORS Sibson D.R. and Gross J.  
 TITLE HUMAN NUCLEIC ACID FRAGMENTS, ISOLATED FROM BRAIN ADRENAL TISSUE,  
 PLACENTA OR BONE MARROW  
 JOURNAL Patent: WO 9401548-A 149 20-JAN-1994.  
 MEDICAL RES COUNCIL (GB); SIBSON DAVID ROSS (GB)  
 FEATURES  
 source 1..204  
 /organism="unidentified"  
 /db\_xref="taxon:32644"

BASE COUNT 58 a 18 c 41 g 81 t 6 others  
 ORIGIN

Query Match 5.7%; Score 116; DB 9; Length 204;  
 Best Local Similarity 100.0%; Pred. No. 6.4e-55;  
 Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1145 attgaatgcttgcgcgaattagaagaagg99gttgcgaattgctgttcttga 1204  
 |||||||  
 DB 16 ATTGAATGCTTACCTGGCCAGATTGAAAGGGGCTTGGAATGGCTGTTGTAAA 75  
 QY 1205 atatacttttgctgtcttgaagtagatagatattacttataataaaaaa 1260  
 |||||||  
 DB 76 ATATATCTTTTACTGTGCTTTAAAGTAGATATACCTTACATTTATATAAAAAA 131

## RESULT 12

LOCUS A77442 204 bp DNA PAT 19-OCT-1999  
 DEFINITION Sequence 149 from Patent EP0587279.  
 ACCESSION A77442  
 VERSION A77442.1 GI:6089107  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.

REFERENCE 1 (bases 1 to 204)  
 AUTHORS Sibson D.R. and Hadfield K.M.  
 TITLE HUMAN NUCLEIC ACID FRAGMENTS ISOLATED FROM BRAIN, ADRENAL TISSUE,  
 PLACENTA OR BONE MARROW AND THEIR USE  
 JOURNAL Patent: EP 0587279-A 149 16-MAR-1994.  
 MEDICAL RES COUNCIL (GB)  
 FEATURES Location/Qualifiers

source 1..204  
 BASE COUNT 58 a 18 c 41 g 81 t 6 others  
 ORIGIN

Query Match 5.7%; Score 116; DB 9; Length 204;  
 Best Local Similarity 100.0%; Pred. No. 6.4e-55;  
 Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1145 attgaatgcttgcgcgaattagaagaagg99gttgcgaattgctgttcttga 1204  
 |||||||  
 DB 16 ATTGAATGCTTACCTGGCCAGATTGAAAGGGGCTTGGAATGGCTGTTGTAAA 75  
 QY 1205 atatacttttgctgtcttgaagtagatagatattacttataataaaaaa 1260  
 |||||||  
 DB 76 ATATATCTTTTACTGTGCTTTAAAGTAGATATACCTTACATTTATATAAAAAA 131

## RESULT 13

LOCUS AL356315 158239 bp DNA HTG 22-MAR-2001  
 DEFINITION Homo sapiens chromosome 1 clone RP11-332P5, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*, 6 unordered pieces.  
 ACCESSION AL356315  
 VERSION AL356315.8 GI:13445377  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 158239)  
 AUTHORS Plumb B.  
 TITLE Direct Submission  
 JOURNAL Submitted (21-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 requests: clonerequest@sanger.ac.uk  
 On Mar 24, 2001 this sequence version replaced gi:13443378.

## COMMENT

Center: Sanger Centre  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquerry@sanger.ac.uk  
 Project Information  
 Center Project name: BA332P5  
 ----- Summary Statistics -----  
 Assembly program: XGAP4; version 4.5  
 Sequencing vector: plasmid; 108752; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Consensus quality: 156356 bases at least Q40  
 Consensus quality: 156868 bases at least Q30  
 Consensus quality: 157303 bases at least Q20  
 Insert size: 157739; sum-of-continigs  
 Insert size: 143450; 18.0% error; agarose-fp  
 Quality coverage: 6.86x in Q20 bases; sum-of-continigs Quality  
 coverage: 7.61x in Q20 bases; agarose-fp  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 6 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 22704: contig of 22704 bp in length  
 \* 22705 22804: gap of 100 bp  
 \* 22805 26580: contig of 3776 bp in length  
 \* 26581 26680: gap of 100 bp  
 \* 26681 53455: contig of 26775 bp in length  
 \* 53456 53555: gap of 100 bp  
 \* 53556 58106: contig of 4551 bp in length

```

* 58107 58206: gap of 100 bp
* 58207 78538: contig of 20332 bp in length
* 78539 78638: gap of 100 bp
* 78639 158239: contig of 79601 bp in length.
FEATURES
    source
        1..158239
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="1"
            /clone="RP11-332P5"
            /clone.lib="RPC1-11.2"
        1..22704
            /note="assembly_fragment:00466
            fragment_chain:1
            clone_end:5P6
            vector_side:left"
        22805..26580
            /note="assembly_fragment:00415
            fragment_chain:1"
        26681..53455
            /note="assembly_fragment:01627
            fragment_chain:1"
        53556..58106
            /note="assembly_fragment:01215
            fragment_chain:1"
        58207..78538
            /note="assembly_fragment:02132
            fragment_chain:2"
        78639..158239
            /note="assembly_fragment:00150
            fragment_chain:2
            clone_end:"7"
            vector_side:right"
BASE COUNT  50161 a 28369 c 28170 g 51037 t 502 others
ORIGIN
Query Match      3.8%; Score 78; DB 80; Length 158239;
Best Local Similarity 100.0%; Pred. No. 1,1e-32;
Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 692 gtcaaacgacgaagcagatcatctccgattatattatctcgtgagcgtgaggtt 751
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 48267 GTCATAGCATCTAGCAAGATCATCTCTCGATTATATTCACGAGCTGCAAGTT 48326
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 752 tggatgaattgggaac 769
    ||||||||||||||||||
Db 48327 TGGATGAATTGGGAAC 48344
    ||||||||||||||||||
RESULT 14
AL356381.c      165617 bp      DNA      HNG      20-JAN-2001
LOCUS           Homo sapiens chromosome 1 clone RP11-294A4, *** SOURCE:NCING IN
DEFINITION      PROGRESS ***; 17 unordered pieces.
ACCESSION       AL356381.7 GI:10039894
VERSION         HNG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS        human.
SOURCE          Homo sapiens
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE       1 (bases 1 to 165617)
AUTHORS        Plumb,B.
TITLE           Direct Submission
JOURNAL         Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
                CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
                requests: clonerequest@sanger.ac.uk
                on Sep 9, 2000 this sequence version replaced gi:9797513.
COMMENT         ----- Genome Center
                Center: Sanger Centre
                Center code: SC
                Web site: http://www.sanger.ac.uk

```

```

Contact: humquerry@sanger.ac.uk
----- Project Information
Center Project name: BA294A4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 159698 bases at least 040
Consensus quality: 161504 bases at least 030
Consensus quality: 162803 bases at least 020
Insert size: 174497; 3.2% error; agarose-fp
Quality coverage: 4.29x in Q20 bases; sum-of-contigs Quality
coverage: 4.21x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1      8110: contig of 8110 bp in length
*      8111 8210: gap of 100 bp
*      8211 13854: contig of 5644 bp in length
*      13855 13954: gap of 100 bp
*      13955 46993: contig of 33039 bp in length
*      46994 47093: gap of 100 bp
*      47094 59514: contig of 12421 bp in length
*      59515 59614: gap of 100 bp
*      59615 86817: contig of 27203 bp in length
*      86818 86917: gap of 100 bp
*      86918 90065: contig of 3148 bp in length
*      90066 90165: gap of 100 bp
*      90166 101858: contig of 11693 bp in length
*      101859 101958: gap of 100 bp
*      101959 105470: contig of 3512 bp in length
*      105471 105570: gap of 100 bp
*      105571 109111: contig of 3541 bp in length
*      109112 109211: gap of 100 bp
*      109212 113589: contig of 4378 bp in length
*      113590 113689: gap of 100 bp
*      113690 124809: contig of 11120 bp in length
*      124810 124909: gap of 100 bp
*      124910 128323: contig of 3414 bp in length
*      128324 128423: gap of 100 bp
*      128424 140194: contig of 11771 bp in length
*      140195 140294: gap of 100 bp
*      140295 149122: contig of 8828 bp in length
*      149123 149222: gap of 100 bp
*      149223 154697: contig of 5475 bp in length
*      154698 154797: gap of 100 bp
*      154798 157559: contig of 2762 bp in length
*      157560 157659: gap of 100 bp
*      157660 165617: contig of 7958 bp in length.
FEATURES
    source
        1..165617
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="1"
            /clone="RP11-294A4"
            /clone.lib="RPC1-11.2"
        1..8110
            /note="assembly_fragment:01010
            fragment_chain:1"
        8211..13854
            /note="assembly_fragment:00790
            fragment_chain:1"
        13955..46993
            /note="assembly_fragment:00399
            fragment_chain:1"
        47094..59514
            /note="assembly_fragment:00399
            fragment_chain:1"
        59515..86817
            /note="assembly_fragment:00399
            fragment_chain:1"
        86818..90065
            /note="assembly_fragment:00399
            fragment_chain:1"
        90066..90165
            /note="assembly_fragment:00399
            fragment_chain:1"
        90166..101858
            /note="assembly_fragment:00399
            fragment_chain:1"
        101859..105470
            /note="assembly_fragment:00399
            fragment_chain:1"
        105471..109111
            /note="assembly_fragment:00399
            fragment_chain:1"
        109112..109211
            /note="assembly_fragment:00399
            fragment_chain:1"
        109212..113589
            /note="assembly_fragment:00399
            fragment_chain:1"
        113590..124809
            /note="assembly_fragment:00399
            fragment_chain:1"
        124810..124909
            /note="assembly_fragment:00399
            fragment_chain:1"
        124910..128323
            /note="assembly_fragment:00399
            fragment_chain:1"
        128324..140194
            /note="assembly_fragment:00399
            fragment_chain:1"
        140195..140294
            /note="assembly_fragment:00399
            fragment_chain:1"
        140295..149122
            /note="assembly_fragment:00399
            fragment_chain:1"
        149123..149222
            /note="assembly_fragment:00399
            fragment_chain:1"
        149223..154697
            /note="assembly_fragment:00399
            fragment_chain:1"
        154698..154797
            /note="assembly_fragment:00399
            fragment_chain:1"
        154798..157559
            /note="assembly_fragment:00399
            fragment_chain:1"
        157560..157659
            /note="assembly_fragment:00399
            fragment_chain:1"
        157660..165617
            /note="assembly_fragment:00399
            fragment_chain:1"

```



```
misc_feature      118021..168860
                  /note="assembly_fragment:01179
                  /fragment_chain:2"
misc_feature      168961..177592
                  /note="assembly_fragment:00474
                  /fragment_chain:2"
misc_feature      17693..193973
                  /note="assembly_fragment:00396
                  /fragment_chain:2"
misc_feature      194074..214692
                  /note="assembly_fragment:01410
                  /fragment_chain:2
                  /vector_side:right"
BASE COUNT      69445 a 38568 c 38843 g 66931 t 905 others
ORIGIN

Query Match      3 84: Score 78; DB 81; Length 214692;
Best Local Similarity 100.0%; Pred.No. 1.le-32;
Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      692 gtcataagcatcagcagatcatctccctgattatatactacgagctgcaagtt 751
          |||||||
Db 199311 GTGATTAAGCATCTAGCAAGATCATCTCGATTATATATCTGAGCTGCGAGGTT 199252
          |||||||

QY      752 tggatgaattgggaagc 769
          |||||||
Db 199251 TGGATGAATTGGGAAGC 199234
          |||||||

RESULT 16
AF259074/c
LOCUS      AF259074      72831 bp      DNA      ROD      21-JUN-2000
DEFINITION Mus musculus T-cell receptor alpha locus BAC clone MBAC519 from
            14D1-D2, complete sequence.
ACCESSION  AF259074
VERSION     AF259074.1 GI:8575574
KEYWORDS
SOURCE      house mouse.
            Mus musculus
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE   1 (bases 1 to 72831)
AUTHORS    Lee,I.Y., Wang,K., Lasky,S.R., Dahl,T., Hall,J. and Hood,L.E.
TITLE      Large-Scale Sequence Analysis of the Mouse T-Cell Receptor
            Alpha/Delta Locus
JOURNAL    Unpublished
REFERENCE   2 (bases 1 to 72831)
AUTHORS    Lee,I.Y., Wang,K., Lasky,S.R., Dahl,T., Hall,J. and Hood,L.E.
TITLE      Submitted (21-APR-2000) Department of Molecular Biotechnology, Box
            357730 University of Washington, Seattle, WA 98195, USA
COMMENT     Interspersed Repeats were identified with RepeatMasker (available
            from http://ftp.genome.washington.edu/RM/RepeatMasker.html) Simple
            sequence repeats were identified with sputnik (available from
            http://serac.mbt.washington.edu/chrisa/software/sputnik.html).
FEATURES             Location/Qualifiers
     source          1..72831
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="14"
                     /map="14D1-D2"
                     /clone="MBAC519; RG:298R19"
                     /cell_line="CJ7"
                     /clone_id="Research Genetics Mouse BAC Library"
                     /note="T-cell receptor alpha locus; flanking sequences are
                     UW:MBAC1058 (Genbank Accession Number AF259072) 5' and
                     UW:MBAC01 (Genbank Accession Number AF259071) 3'"
     repeat_region   /rpt_family="L1_MM"
                     complement(1..3813)
                     /rpt_family="L1_MM"
                     complement(3794..9677)
                     /rpt_family="L1_MM"

repeat_region     complement(9585..10193)
                  /rpt_family="L1F_5end"
repeat_region     complement(9922..10405)
                  /rpt_family="L1F_5end"
repeat_region     complement(10134..10617)
                  /rpt_family="L1F_5end"
repeat_region     complement(10346..10666)
                  /rpt_family="L1F_5end"
repeat_region     complement(10688..11092)
                  /rpt_family="L1M2_orf2"
repeat_region     complement(11099..11689)
                  /rpt_family="L1_RN"
repeat_region     11945..11972
                  /rpt_family="CAU)n"
repeat_region     12294..12490
                  /rpt_family="RLTR13D"
repeat_region     13228..13278
                  /rpt_family="TGC)n"
repeat_region     13291..13464
                  /rpt_family="TGC)n"
repeat_region     14374..15028
                  /rpt_family="RLTR13C"
repeat_region     14515..15028
                  /rpt_family="RLTR13D"
repeat_region     15157..15862
                  /rpt_family="L1_MM"
repeat_region     15873..16356
                  /rpt_family="L1F_5end"
repeat_region     16085..16568
                  /rpt_family="L1F_5end"
repeat_region     16509..16992
                  /rpt_family="L1F_5end"
repeat_region     16933..17541
                  /rpt_family="L1F_5end"
repeat_region     17449..23342
                  /rpt_family="L1_MM"
repeat_region     23348..23928
                  /rpt_family="L1_MM"
repeat_region     23365..24582
                  /rpt_family="L1_RN"
repeat_region     24584..24701
                  /rpt_family="L1_MM"
repeat_region     24599..26220
                  /rpt_family="L1_MM"
repeat_region     complement(26217..27218)
                  /rpt_family="L1_MM"
repeat_region     complement(27118..27357)
                  /rpt_family="L1F_5end"
repeat_region     27881..28092
                  /rpt_family="B3"
repeat_region     complement(28543..28587)
                  /rpt_family="CA)n"
repeat_region     complement(28929..28969)
                  /rpt_family="CA)n"
repeat_region     28938..28969
                  /rpt_family="CT)x15
repeat_region     29384..29531
                  /rpt_family="MTD"
repeat_region     29556..29866
                  /rpt_family="L1_RN"
repeat_region     29875..30280
                  /rpt_family="RLTR13C"
repeat_region     30282..31643
                  /rpt_family="L1_MM"
repeat_region     31674..32181
                  /rpt_family="RLTR13C"
repeat_region     32186..34760
                  /rpt_family="L1_MM"
repeat_region     34747..36147
                  /rpt_family="L1_RN"
repeat_region     34929..34953
                  /rpt_family="AAAC)x6
repeat_region     34930..34953
```



```

repeat_region /rpt_family="(CAAA)n"
36156..36306
repeat_region /rpt_family="Lx3"
36337..36557
repeat_region /rpt_family="RMR13A"
36561..37025
repeat_region /rpt_family="RMR13A"
37029..37841
repeat_region /rpt_family="Lx4"
37821..38430
repeat_region /rpt_family="Lx4"
complement(38431..38476)
repeat_region /rpt_family="AT_rich"
39113..39169
repeat_region /rpt_family="(CA)x28"
39114..39169
repeat_region /rpt_family="(CA)n"
complement(39309..40536)
repeat_region /rpt_family="L1_MM"
complement(40537..40629)
repeat_region /rpt_family="Lx5"
40636..40687
repeat_region /rpt_family="RMR6A"
40688..41215
repeat_region /rpt_family="RMR6A"
complement(41325..41441)
repeat_region /rpt_family="RMR1B"
41452..41573
repeat_region /rpt_family="ORRA3"
42854..42984
repeat_region /rpt_family="B1-F"
42986..43037
repeat_region /rpt_family="(CAAA)n"
complement(44550..44693)
repeat_region /rpt_family="RMR1B"
complement(44699..44893)
repeat_region /rpt_family="RMR1B"
44997..45152
repeat_region /rpt_family="MERVL"
45066..45164
repeat_region /rpt_family="MERVL"
45202..46018
repeat_region /rpt_family="MERVL"
complement(46029..46760)
repeat_region /rpt_family="L1_MM"
46762..46985
repeat_region /rpt_family="MERVL"
complement(46986..47171)
repeat_region /rpt_family="ORRA1"
complement(47015..47353)
repeat_region /rpt_family="ORRIC"
47201..47258
repeat_region /rpt_family="(AG)x28"
47202..47258
repeat_region /rpt_family="(GA)n"
complement(47350..47483)
repeat_region /rpt_family="RMR1B"
complement(47643..48403)
repeat_region /rpt_family="RMR1B"
complement(48459..48558)
repeat_region /rpt_family="ORRIC"
48573..48751
repeat_region /rpt_family="MERVL"
48611..48833
repeat_region /rpt_family="MERVL"
48940..49259
repeat_region /rpt_family="L1_MM"
49288..49549
repeat_region /rpt_family="MERVL"
complement(49570..49703)
repeat_region /rpt_family="(TA)n"
49751..49830
repeat_region /rpt_family="B1-F"

```

```

repeat_region complement(49884..50083)
repeat_region /rpt_family="LIME2"
repeat_region complement(49885..50110)
repeat_region /rpt_family="LIME1"
51858..51878
repeat_region /rpt_family="AT_rich"
52211..52307
repeat_region /rpt_family="PBID9"
52224..52294
repeat_region /rpt_family="B1_MM"
52319..52422
repeat_region /rpt_family="L1M2_orf2"
52424..52487
repeat_region /rpt_family="B1-F"

```

```

Query Match 14% Score 29; DB 94; Length 72831;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 421 gtgcgaattccattccctatttc 449
Db 43798 GTTGCAAATTCATTCACCTTATTTTC 43770
|||||

```

```

RESULT 17
AL354916 153519 bp DNA HTG 10-APR-2001
LOCUS Homo sapiens chromosome 10 clone RP11-26K22, *** SEQUENCING IN
DEFINITION PROGRESS ***; 16 unordered pieces.
ACCESSION AL354916
VERSION AL354916.11 GI:13619224
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

```

```

REFERENCE 1
AUTHORS Submitted (09-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
TITLE CB10 15A, UK. E-mail enquiries: humquery@sanger.ac.uk
JOURNAL Requests: clonequests@sanger.ac.uk
1 (bases 1 to 153519)

```

```

COMMENT On Apr 12, 2001 this sequence version replaced g1:11229129.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba26K22
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 147377 bases at least Q40
Consensus quality: 149649 bases at least Q20
Insert size: 152019; sum-of-contigs
Insert size: 170587; 5.9% error; agarose-fp
Quality coverage: 3.81x in Q20 bases; sum-of-contigs quality
coverage: 3.47x in Q20 bases; agarose-fp
-----

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

* 1 11258: contig of 11258 bp in length
* 11259 11358: gap of 100 bp
* 11359 16689: contig of 5331 bp in length

```

```

* 16690 16789: gap of 100 bp
* 16790 19081: contig of 2292 bp in length
* 19082 19181: gap of 100 bp
* 19182 28722: contig of 9541 bp in length
* 28723 28822: gap of 100 bp
* 28823 34712: contig of 5890 bp in length
* 34713 34812: gap of 100 bp
* 34813 42919: contig of 8107 bp in length
* 42920 43019: gap of 100 bp
* 43020 50168: contig of 7149 bp in length
* 50169 50268: gap of 100 bp
* 50269 60388: contig of 10120 bp in length
* 60389 60488: gap of 100 bp
* 60489 67782: contig of 7294 bp in length
* 67783 67882: gap of 100 bp
* 67883 74589: contig of 6707 bp in length
* 74590 74689: gap of 100 bp
* 74690 95906: contig of 21217 bp in length
* 95907 96006: gap of 100 bp
* 96007 119368: contig of 23362 bp in length
* 119369 119468: gap of 100 bp
* 119469 126722: contig of 7254 bp in length
* 126723 126822: gap of 100 bp
* 126823 139086: contig of 12264 bp in length
* 139087 139186: gap of 100 bp
* 139187 141602: contig of 2416 bp in length
* 141603 141702: gap of 100 bp
* 141703 153519: contig of 11817 bp in length.

```

## FEATURES

```

source
1..153519
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="10"
  /clone="RP11-26k22"
  /clone_id="RPCL11.1"
  1..11258
misc_feature
  /note="assembly_fragment:00631
  fragment_chain:1"
  11359..16689
misc_feature
  /note="assembly_fragment:00946
  fragment_chain:1"
  16790..19081
misc_feature
  /note="assembly_fragment:00025
  fragment_chain:1"
  19182..28722
misc_feature
  /note="assembly_fragment:00181
  fragment_chain:1"
  28823..34712
misc_feature
  /note="assembly_fragment:01377
  fragment_chain:1"
  34813..42919
misc_feature
  /note="assembly_fragment:00226
  fragment_chain:1"
  43020..50168
misc_feature
  /note="assembly_fragment:01546
  fragment_chain:1"
  50269..60388
misc_feature
  /note="assembly_fragment:00124
  fragment_chain:1"
  60489..67782
misc_feature
  /note="assembly_fragment:00414
  fragment_chain:1"
  67883..74589
misc_feature
  /note="assembly_fragment:00151
  fragment_chain:2"
  74690..95906
misc_feature
  /note="assembly_fragment:01117
  fragment_chain:2"
  96007..119368
misc_feature
  /note="assembly_fragment:00762
  fragment_chain:3"
  119469..126722
  /note="assembly_fragment:01744

```

```

misc_feature      fragment_chain:3"
126823..139086
  /note="assembly_fragment:00202"
misc_feature      139187..141602
  /note="assembly_fragment:00844"
misc_feature      141703..153519
  /note="assembly_fragment:01714"
BASE COUNT      47293 a 28241 c 28819 g 47661 t 1505 others
ORIGIN

```

```

Query Match      13%, Score 27; DB 79; Length 153519;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1245 acattataaaaaaacatttg 1271
Db 92911 ACATTATVAAAAAAATCAATTTCG 92937
|||||

```

```

RESULT 18
AL512657
LOCUS
DEFINITION
  Homo sapiens chromosome 10 clone RP11-265H3, *** SEQUENCING IN
  PROGRESS ***, 5 unordered pieces.
ACCESSION
  AL512657
VERSION
  AL512657.5 GI:13374822
KEYWORDS
  HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 154776)
Sims.S.
  Direct Submission
  Submitted (12-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
  CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
  requests: clonerequest@sanger.ac.uk
  On Mar 17, 2001 this sequence version replaced gi.13374028.
  ----- Genome Center
  Center: Sanger Centre
  Center code: SC
  Web site: http://www.sanger.ac.uk
  Contact: humquerry@sanger.ac.uk
  ----- Project Information
  Center project name: BA265H3
  ----- Summary Statistics
  Assembly program: XGAP4; version 4.5
  Sequencing vector: plasmid; L08752; 100% of reads
  Chemistry: Dye-terminator Big Dye; 100% of reads
  Consensus quality: 153759 bases at least Q40
  Consensus quality: 154025 bases at least Q30
  Consensus quality: 154222 bases at least Q20
  Insert size: 154376; sum-of-ctrligs
  Insert size: 156931; 3.4% error; agarose-fp
  Quality coverage: 7.62x in Q20 bases; sum-of-ctrligs Quality
  coverage: 7.49x in Q20 bases; agarose-fp
  -----

```

## COMMENT

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 53336: contig of 53336 bp in length
* 53337 53436: gap of 100 bp
* 53437 77952: contig of 24516 bp in length
* 77953 78052: gap of 100 bp
* 78053 86355: contig of 8303 bp in length
* 86356 86455: gap of 100 bp
* 86456 97941: contig of 11486 bp in length

```

```
*      *      97942 98041: gap of 100 bp
*      *      98042 154776: contig of 56735 bp in length.
FEATURES
      *      1. 154776
      source      /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="10"
                  /clone_id="RP11-265H3"
                  /clone_lib="RPC1-11.1"
                  1. 53336
misc_feature      /note="assembly: fragment:02560
                  fragment_chain:1
                  clone_end:SP6
                  vector_side:left"
misc_feature      53437..77952
                  /note="assembly: fragment:02637
                  fragment_chain:1"
misc_feature      78053..86355
                  /note="assembly: fragment:01421
                  fragment_chain:1"
misc_feature      86456..97941
                  /note="assembly: fragment:01318
                  fragment_chain:1"
misc_feature      96042..154776
                  /note="assembly: fragment:02753
                  fragment_chain:1
                  clone_end:77
                  vector_side:right"
BASE COUNT      50002 a 28250 c 28911 g 47213 t 400 others
ORIGIN
Query Match      1.3%; Score 27; DB 81; Length 154776;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1245 acattataaaaaaaccaatttg 1271
    |||||||
Db 5428 ACATTATATATATATATATATTTTG 5454

RESULT 19
AC025817 162696 bp DNA HTG 07-JUL-2000
LOCUS      Homo sapiens chromosome 3 clone RP11-14301, WORKING DRAFT SEQUENCE,
DEFINITION 27 unordered pieces.
ACCESSION  AC025817.2 GI:7637787
VERSION     HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS    human.
SOURCE      Homo sapiens
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 162696)
AUTHORS    Waterston,R.H.
TITLE       The sequence of Homo sapiens clone
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 162696)
AUTHORS    Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (15-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Apr 23, 2000 this sequence version replaced gl:7243897.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0143001
----- Summary Statistics -----
Sequencing vector: M13; 100%
```

```
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 148785 bases at least Q40
Consensus quality: 152012 bases at least Q30
Consensus quality: 153717 bases at least Q20
Insert size: 175000; agarose-fp
Insert size: 160096; sum-of-contigs
Quality coverage: 3.80 in Q20 bases; agarose-fp
Quality coverage: 4.29 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1573: contig of 1573 bp in length
* 1574 1673: gap of unknown length
* 1674 3146: contig of 1473 bp in length
* 3147 3246: gap of unknown length
* 3247 4594: contig of 1348 bp in length
* 4595 4694: gap of unknown length
* 4695 7289: contig of 2594 bp in length
* 7289 7389: gap of unknown length
* 7389 8734: contig of 1345 bp in length
* 8734 8833: gap of unknown length
* 8833 12338: contig of 3505 bp in length
* 12339 12438: gap of unknown length
* 12439 15581: contig of 3143 bp in length
* 15582 15681: gap of unknown length
* 15682 18712: contig of 3031 bp in length
* 18713 18812: gap of unknown length
* 18813 21943: contig of 3131 bp in length
* 21944 22043: gap of unknown length
* 22044 24432: contig of 2389 bp in length
* 24433 24532: gap of unknown length
* 24533 26675: contig of 4143 bp in length
* 26676 28775: gap of unknown length
* 28776 33292: contig of 4517 bp in length
* 33293 33392: gap of unknown length
* 33393 37585: contig of 4193 bp in length
* 37586 37685: gap of unknown length
* 37686 42468: contig of 4783 bp in length
* 42469 42568: gap of unknown length
* 42569 47416: contig of 4848 bp in length
* 47417 47516: gap of unknown length
* 47517 52817: contig of 5301 bp in length
* 52818 52917: gap of unknown length
* 52918 59878: contig of 6961 bp in length
* 59879 59978: gap of unknown length
* 59979 66303: contig of 6325 bp in length
* 66304 66403: gap of unknown length
* 66404 72170: contig of 5767 bp in length
* 72171 72270: gap of unknown length
* 72271 77598: contig of 5328 bp in length
* 77599 77698: gap of unknown length
* 77699 87929: contig of 10231 bp in length
* 87930 88029: gap of unknown length
* 88030 95226: contig of 7197 bp in length
* 95227 95327: gap of unknown length
* 95327 102793: contig of 7467 bp in length
* 102794 102893: gap of unknown length
* 102894 112129: contig of 9236 bp in length
* 112130 112229: gap of unknown length
* 112230 125888: contig of 13659 bp in length
* 125889 125988: gap of unknown length
* 125989 143362: contig of 17374 bp in length
* 143363 143462: gap of unknown length
* 143463 162696: contig of 19234 bp in length.
```

FEATURES  
source

Location/Qualifiers  
1. .162696  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/clone="RP11-14301"  
1. .1573  
/note="assembly\_name:Contig5"  
1674. .3146  
/note="assembly\_name:Contig7"  
3247. .4594  
/note="assembly\_name:Contig8"  
4695. .7288  
/note="assembly\_name:Contig9"  
7389. .8733  
/note="assembly\_name:Contig10  
clone\_end:T7  
vector\_side:left"  
8834. .12338  
/note="assembly\_name:Contig11"  
12439. .15581  
/note="assembly\_name:Contig12"  
15682. .18712  
/note="assembly\_name:Contig13"  
18813. .21943  
/note="assembly\_name:Contig14"  
22044. .24432  
/note="assembly\_name:Contig15  
clone\_end:SP6  
vector\_side:right"  
24533. .28675  
/note="assembly\_name:Contig16"  
28776. .33292  
/note="assembly\_name:Contig17"  
33393. .37585  
/note="assembly\_name:Contig18"  
37686. .42468  
/note="assembly\_name:Contig19"  
42569. .47416  
/note="assembly\_name:Contig20"  
47517. .52817  
/note="assembly\_name:Contig21"  
52918. .59878  
/note="assembly\_name:Contig22"  
59979. .66303  
/note="assembly\_name:Contig23"  
66404. .72170  
/note="assembly\_name:Contig24"  
72271. .77598  
/note="assembly\_name:Contig25"  
77699. .87929  
/note="assembly\_name:Contig26"  
88030. .95226  
/note="assembly\_name:Contig27"  
95327. .102793  
/note="assembly\_name:Contig28"  
102894. .112129  
/note="assembly\_name:Contig29"  
112230. .125888  
/note="assembly\_name:Contig30"  
125989. .143362  
/note="assembly\_name:Contig31"  
143463. .162696  
/note="assembly\_name:Contig32"  
BASE COUNT 49514 a 31064 c 29890 g 49621 t 2607 others  
ORIGIN

DB 96952 CATTATTTCAGTACTTTTC 96974

|||||

RESULT 20  
AC026127  
LOCUS  
DEFINITION Homo sapiens chromosome 3 clone RP11-89F18, WORKING DRAFT SEQUENCE,  
11 unordered pieces.  
ACCESSION  
AC026127.14 GI:12831262  
VERSION  
KEYWORDS  
SOURCE HTGS\_PHASE1; HTGS\_DRAFT.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 171993)

REFERENCE  
AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbara,J.,  
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,  
Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,  
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
Dem,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C.,  
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,  
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,  
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,  
Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,  
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,  
Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,  
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,  
Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korah,J.,  
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,  
Lewisl,I., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,  
Louieged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,  
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,  
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,  
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,  
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkwo,S.,  
Ogun,M., Okwoudu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,  
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,  
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojupokan,I., Rolfe,M.,  
Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shoshari,N.,  
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,  
Stinson,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,  
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,  
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R.,  
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,  
Washington,S., Williams,G., Williams,A., Wlezy,R., Wooden,S.,  
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.  
and Gibbs,R.  
Direct Submission  
Unpublished  
2 (bases 1 to 171993)  
Worley,K.C.  
Direct Submission  
Submitted (19-MAR-2000) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Feb 15, 2001 this sequence version replaced gi:9664953.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: HAOO  
Center clone name: RP11-89F18

Query Match 1.1%; Score 23; DB 69; Length 162696;  
Best Local Similarity 100.0%; Pred. No. 0.56;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1870 catttatttcagtaacttttc 1892

## ----- Summary Statistics

Sequencing vector: M13; L08821  
Chemistry: Dye-Primer Bodypy: 4% of reads  
Chemistry: Dye-terminator Big Dye: 96% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 15369 bases at least Q40  
Consensus quality: 16268 bases at least Q30  
Consensus quality: 16628 bases at least Q20  
Estimated insert size: 167989; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-gel estimation  
Quality coverage: 4.3x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank/draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 11 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 36507: contig of 36507 bp in length  
\* 36508 36607: gap of unknown length  
\* 36608 57065: contig of 20458 bp in length  
\* 57066 57165: gap of unknown length  
\* 57166 78181: contig of 21016 bp in length  
\* 78182 98347: gap of unknown length  
\* 98348 98347: contig of 19966 bp in length  
\* 98348 115329: contig of 16882 bp in length  
\* 115330 115329: gap of unknown length  
\* 115330 130420: contig of 15091 bp in length  
\* 130421 130520: gap of unknown length  
\* 130521 141163: contig of 10643 bp in length  
\* 141164 141263: gap of unknown length  
\* 141264 150389: contig of 9126 bp in length  
\* 150390 150489: gap of unknown length  
\* 150490 158813: contig of 8324 bp in length  
\* 158814 158914: gap of unknown length  
\* 158915 165814: contig of 6901 bp in length  
\* 165815 165814: gap of unknown length  
\* 165915 171993: contig of 6079 bp in length.  
Location/Qualifiers

FEATURES  
source  
1. 171993  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/clone="RP11-89F18"

BASE COUNT 53426 a 32407 c 32843 g 52491 t 1026 others  
ORIGIN

Query Match 1.1%; Score 23; DB 69; Length 171993;  
Best Local Similarity 100.0%; Pred. No. 0.56;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1870 gattttatttcagtaatttc 1892  
|||||  
Db 51044 CATTATTTATTCAGTAATTTC 51066

RESULT 21  
AC023818/c AC023818 181413 bp DNA HTG 11-AUG-2000  
LOCUS Homo sapiens chromosome 16 clone CTD-2600H12, WORKING DRAFT  
DEFINITION  
SEQUENCE, 10 ordered pieces.  
AC023818  
AC023818.3 GI:8576137  
VERSION  
KEYWORDS HTG: HTGS\_PHASE2; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## REFERENCE

1 (bases 1 to 181413)  
DOE Joint Genome Institute.  
Sequencing of Human Chromosome 16  
Unpublished  
JOURNAL  
2 (bases 1 to 181413)  
DOE Joint Genome Institute.  
Direct Submission  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Submitted (18-FEB-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Jun 21, 2000 this sequence version replaced g1:7211920.  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>  
-----

Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
Genbank flat file format but are available as part  
of this entry's ASN.1 file.

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 10 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.

\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

1 14962: contig of 14962 bp in length  
\* 14963 15062: gap of unknown length  
\* 15063 16882: contig of 1820 bp in length  
\* 16883 16982: gap of unknown length  
\* 16983 24654: contig of 7672 bp in length  
\* 24655 24754: gap of unknown length  
\* 24755 98161: contig of 73407 bp in length  
\* 98162 98261: gap of unknown length  
\* 98262 104627: contig of 6366 bp in length  
\* 104628 104727: gap of unknown length  
\* 104728 124624: contig of 19897 bp in length  
\* 124625 124724: gap of unknown length  
\* 124725 133394: contig of 8670 bp in length  
\* 133395 133494: gap of unknown length  
\* 133495 139249: contig of 5735 bp in length  
\* 139250 139249: gap of unknown length  
\* 139250 168158: contig of 28809 bp in length  
\* 168159 168258: gap of unknown length  
\* 168259 181413: contig of 13155 bp in length.  
Location/Qualifiers

FEATURES  
source  
1. 181413  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="CTD-2600H12"

BASE COUNT 51370 a 37917 c 38106 g 53071 t 949 others  
ORIGIN

Query Match 1.1%; Score 23; DB 68; Length 181413;  
Best Local Similarity 100.0%; Pred. No. 0.57;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1288 tgcctgacatcttctagact 1310  
|||||  
Db 80363 TGCCTGATGATCTTTCTAGACT 80341

RESULT 22  
CEY102A5A 11230 bp DNA INV 25-OCT-2000  
LOCUS Caenorhabditis elegans cosmid Y102A5A, complete sequence.  
DEFINITION  
ACCESSION AL023822.299711  
VERSION AL023822.1 GI:3646744  
KEYWORDS  
SOURCE HTG.  
ORGANISM  
Caenorhabditis elegans.  
Caenorhabditis elegans.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
Rhabditioidea; Rhabditidae; Pelodierinae; Caenorhabditis.  
1 (bases 1 to 11230)  
none.  
Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium  
Science 282 (5396), 2012-2018 (1998)  
99069613  
The C. elegans Sequencing Consortium.  
Erratum: [[published errata appear in Science 1999 Jan  
1, 283(5398):35 and 1999 Mar 26;283(5410):2103 and 1999 Sep  
3, 283(5433):1493]]  
2 (bases 1 to 11230)  
Gardner, A.E.  
Direct Submission  
Submitted (09-JUN-1998) Nematode Sequencing Project, Sanger Centre,  
Hinxton, Cambridge CB10 1RO, England and Department of Genetics,  
Washington University, St. Louis, MO 63110, USA. E-mail:  
jesesanger.ac.uk or twenematode.wustl.edu  
On Sep 24, 1998 this sequence version replaced gi:3217805.  
Coding sequences below are predicted from computer analysis, using  
predictions from GeneFINDER (P. Green, U. Washington), and other  
available information.  
Current sequence finishing criteria for the C. elegans genome  
sequencing consortium are that all bases are either sequenced  
unambiguously on both strands, or on a single strand with both a  
dye primer and dye terminator reaction, from distinct subclones.  
Exceptions are indicated by an explicit note.  
IMPORTANT: This sequence is not the entire insert of clone Y102A5A.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we arrange for a small overlap between  
neighbouring subclones.  
The true left end of clone Y102A5 is at 1 in this sequence. The  
true left end of clone T20B3 is at 11127 in this sequence. The  
start of this sequence (1..104) overlaps with the end of sequence  
AL021572.  
The end of this sequence (11127..11230) overlaps with the start of  
sequence Z81593.  
For a graphical representation of this sequence and its analysis  
see: [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=Y102A5A)  
[name=Y102A5A](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=Y102A5A).

FEATURES  
source  
1..11230  
Location/Qualifiers  
/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"  
/chromosome="V"  
/clone="Y102A5A"  
join(2265..2402,2895..3005,3540..3794,4465..4978,  
5796..6679,7348..7524,7941..8094,8593..8684,9399..10036,  
10916..11230,281593.1:105..209,281593.1:983..1352,  
281593.1:2277..2348)  
/gene="Y102A5A.1"  
join(2265..2402,2895..3005,3540..3794,4465..4978,  
5796..6679,7348..7524,7941..8094,8593..8684,9399..10036,  
10916..11230,281593.1:105..209,281593.1:983..1352,  
281593.1:2277..2348)  
/note="CDNA EST YK8H4.3 comes from this gene  
CDNA EST YK8H4.5 comes from this gene  
CDNA EST YK2004.5 comes from this gene  
CDNA EST YK38369.3 comes from this gene  
CDNA EST YK247d1.5 comes from this gene  
CDNA EST YK43098.5 comes from this gene  
CDNA EST YK383c8.5 comes from this gene"

CDS  
gene

CDNA EST YK530b2.5 comes from this gene  
CDNA EST YK563b2.5 comes from this gene  
CDNA EST YK572c8.5 comes from this gene  
/protein\_id="CA19440.1"  
/db\_xref="GI:3880634"  
/db\_xref="SPTREMBL:O9XTJ0"  
/translation="MSAYVGVGLVDMKSNPDKDFPMACNDLMDLOTGTALPEDST  
AKVIRALIKLSDNGEONLAKICGLLAOPKIKTHHLELYVEETLTPVSKAES  
RDIHSILTKAMILNLRPSASSMATTYYKMLPKPYDSLSICAPDDAARDYDLIGC  
VLRGDIVPEMRKGLKAYMDHLYSRSAIRKATGTIGLASVINGELYDELVDL  
LKEAQRSPSSAQNVDLRLVIALSTVARAGSSEKHPKVPPLLYLDIDPGT  
ESEHDDLEASIGLEVLVLRNPOVFAFEVYLOULDALAPNPDGDDDEDM  
EDEDDEDESDDEDEYVWKVRAAKAIEAMISHRSLNLSOKIGPVGRKER  
EETVATEITIVYIALLNLSLIPDLQKAVVADEDSLETDIYVIGRFSNVLK  
SOLATIOSLQDKDVLRLITRKSMKHPKGTGPECIEELSLIRYPGSESLDITP  
AVSNLIDKKAASOGKMYLSIPISNATLNNPRKKNLSPTITMHSISPEPKVS  
AEGALVCKRTIVRLSLACSGNEAKLLVVEKKPMANDTDOERERISISMIL  
AARFQVLRNETPAIDKMTERRIGDMTCVAFRASHVIVEGILFESSAOLSLRHV  
DYVKRIARSLRMTCLNPEVKLMSKSPAGSIPEVELCVLIGMSMLISEDLOITNQAR  
COLTYAFANPPTCVSLHMQPIIDSIIRLISPLIOGLANSLNLFYAIKDPPEKP  
TFESLDSVTPYVDNVALISRAHAMAIASCAVITESTQNLKSRSLAKLAQLOTA  
NMSDSIRLFAMITLIGELGRVDTSPDPVPRKEDLAKFNHHEDLKSAALGAI  
LAVGNLVVYLPFTIEGIRPOKROYLLHLAKREVLVWSSSESTSTDLPRSAVAK  
WGMMAVAGNENGGRSVYACIGRLCSPPESLIPKIKESMSRSDPAGRSVAIAIK  
YVNDERIYDITLQKIGDPLAVRDEDIKVRALVYVLSAHHKPRALVYDLPLD  
LPAYEETKRLKELREVMGPFKHQYDEGLDKCAFECMTLLESQVCKDIDTIFS  
SVMEVGLSDNHDVKLNLITLQRLVNAALPQVQVLRDVECPQLQVLRGRNAVK  
QEVKLEIKLAVIRVYGLKIKLPEVERNPQFLDYNTIKHKELEALANDVLEKSO  
RAVYDDPMETA"

BASE COUNT 3761 a 1935 c 1830 g 3704 t  
ORIGIN

Query Match 1.1%; Score 22; DB 6; Length 11230;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1252 taataaaatcaattgtt 1273  
|||||  
Db 790 TAAAAAAATCAATTGTGT 811

RESULT 23  
CEAH10  
LOCUS 31039 bp DNA INV 03-APR-2001  
DEFINITION Caenorhabditis elegans cosmid AH10, complete sequence.  
ACCESSION Z81027  
VERSION Z81027.1 GI:1894779  
KEYWORDS  
SOURCE HTG.  
ORGANISM  
Caenorhabditis elegans.  
Caenorhabditis elegans.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
Rhabditioidea; Rhabditidae; Pelodierinae; Caenorhabditis.  
1 (bases 1 to 31039)  
none.  
Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium  
Science 282 (5396), 2012-2018 (1998)  
99069613  
The C. elegans Sequencing Consortium.  
2 (bases 1 to 31039)  
McMurray, A.A.  
Direct Submission  
Submitted (21-OCT-1996) Nematode Sequencing Project, Sanger Centre,  
Hinxton, Cambridge CB10 1RO, England and Department of Genetics,  
Washington University, St. Louis, MO 63110, USA. E-mail:  
jesesanger.ac.uk or twenematode.wustl.edu  
On Mar 18, 1997 this sequence version replaced gi:1627523.  
Coding sequences below are predicted from computer analysis, using  
predictions from GeneFINDER (P. Green, U. Washington), and other  
available information.  
Current sequence finishing criteria for the C. elegans genome







2117. .2167)  
/gene="C33H5.7"  
/note="similar to G beta repeats (PROSITE:PS00670)"  
/codon\_start=1  
/evidence=not experimental  
/protein\_id="AA82264.1"  
/db\_xref="GI:1086586"  
/translation="MTKEVEITRESLAKLTIPAKOPNDTTRVNSISYSGKCDHINS  
TDDSMFLYDNLGRLSRVSKKYGCSHVFASDSCALHSGTKYDNTIRLSELDN  
KYIRYQGHDKLVGTINVSPPMDMFLSVADKTTLMOLKTHNCIGLMLNLSKPLTAR  
DPEGLLEAGLDNNITIKLYLDRSPDGSPSSGPLENDNVEWTAMRFSPGKYLMT  
TNSQGLLDVAFGLIKFTLEHONAKIPLAASFTPESSHIVGSSDDYIFEDYDET  
GEILKTIAPNNOTCHIAEFSPHFVAATADTKLTLMCASDELKAAQOS"  
2405. .3597  
/gene="C33H5.6"  
/join(2405. .2470,2517. .2642,2689. .2937,3292. .3597)  
/gene="C33H5.6"  
/note="similar to G beta repeats (PROSITE:PS00670)"  
/codon\_start=1  
/evidence=not experimental  
/protein\_id="AA82265.1"  
/db\_xref="GI:1086587"  
/translation="MYNDPELITVOPAPKALTPSKKNENSKLIDLVASDGNLI  
VSAEDVLYLDLSDSGSKSTECKKYVNCLELDTVYCVHSGNDTTRLLHLOKRS  
YVRYEDHRMVPNFIETNPDKREKFLSSVDGELTMDPTRENTYGLGHTPLIAR  
DPDGLFNAVKSDTIKLEDMRSPDLGPQSFITTKDNDENWIDIESPGKFLILST  
TAGSVKIDAFDGTILHSFSQHKNNVSFP"  
complement(4401. .5343)  
/gene="C33H5.8"  
complement(join(4401. .4670,4720. .5343))  
/gene="C33H5.8"  
/note="similar to mitochondrial precursor proteins import  
receptors"  
/codon\_start=1  
/evidence=not experimental  
/protein\_id="AA82267.1"  
/db\_xref="GI:1086589"  
/translation="MDEEIKIAPRLKBOGNEAFKKKKYHKAMTYSKLEHMPDPIV  
SNRQAGLNDLPIALQIDCTALINDSTPAKAYVRAQAFKALELAEARMTCF  
KISNDPMERQANDLKGGKNNQYDIDPSTRDEYLOSDEKFTKIDPYNTEKEETV  
AEEPNDNIVPKYITIKLPHRPDQVDFGVANMLKRAOSLPLAETFLNLSMDKSEL  
FDELLDVAVALFEKALNHLKGTGTIPHLAAMTISBELSFLLIAFMSERQSTI  
TEILDVYVPEEAMVSRSYF"  
complement(5932. .8744)  
/gene="C33H5.9"  
complement(join(5932. .6520,6568. .6800,7538. .8641,  
8691. .8744))  
/gene="C33H5.9"  
/note="coded for by C. elegans cDNA yk31d12.5; coded for  
by C. elegans cDNA yk31d12.3"  
/codon\_start=1  
/protein\_id="AA82266.1"  
/db\_xref="GI:1086588"  
/translation="MRFEETISLOMLCDOFONKINTLEKMEKKEDYOKLHRLHE  
KNGELDKMOLDHQMAVSTKVYVLDGLDSEVDOPRSRAHQLHQLHDEFLSNOP  
LNSAIFTDPDKLESADLVHRLKYSOELNKDKPANYARICORRYVEDLLIEEVR  
SORDKKMAEVAKTILSEFGISGCVDRIVPLCSITHPGDGEGLADCLQICLRTQOP  
RISALFEPHYVQKLVNLIFGRKKETITARRACKDTHDDEHLDLAYLSLTK  
MKELEKLIHSPDSAFILSTLSDISFORIATYSEELKYLNDQCSNLLQRYFSKRVH  
KROIGGGLHKLKDVAAKLAMNEYGGTEFAEDVAISILQETKAGRGARNOICDEE  
LRRHENIVDLKLYLGEHLDAVEGTLAGISLAESKEPAPYFFSVAKKISVILL  
MKOPEDEIPITIKETIVESPVAKKMOOSLSLESKMSLGLEROLNSIVGVFELSE  
OKKTRPDSQOQIDIRVSPQGLARFLANQVAMELGCGENLEALQSDLATRLKPF  
MHTHCOPTVNSGAVLLICDYGELRTIYSKWRVONALQWESLQALINLVALPQV  
NETAHSLSLENVROLIHPFVYLRTPDFRSIKNFOI"

/evidence=not experimental  
/protein\_id="AA82269.1"  
/db\_xref="GI:1086591"  
/translation="MSDGLTDFRANQIIQYCAEIRELYETIEEYVROPNDHLEFLYD  
LSTFLLELCPEHELCGDIETRFQSSAKKLLSPFLDSVTKARLITRONTEGELIT  
DGRYSOSORHPLFKATIDENLTPEIKQCESRDFYNNRLILNLSKAVYVESLMD  
RLKMHENIOSILSARVDEIGRVKANSDVAILLSASSLLYIQKASMDREBTRSKO  
HPLSIDAPKDRGRTGEMVGVKQESFROOCHTFDROHSONROKSTFDPRYTHEOVL  
EOLHSRDNSGMRGRGRGRGRGRGRG"  
complement(14117. .15827)  
/gene="C33H5.11"  
complement(join(14117. .14203,14257. .14483,15137. .15239,  
15510. .15668,15720. .15827))  
/gene="C33H5.11"  
/note="coded for by C. elegans cDNA cm01a10"  
/codon\_start=1  
/protein\_id="AA82268.1"  
/db\_xref="GI:1086590"  
/translation="MRDVKAKIAASLILSLVYVHSGRRASTIYFPELDTTVIF  
MDVYPAMTLEFVFTVLYETLGTKRSANTERETAREAKELKTSSEFNRSKYSE  
ISRAVSGLDQAEQOQIRKSSVTPPEVRSRTSKSTHDLMTIRGSSNMISASO  
RSRSKSSMGSPDKSQIYWTGEGCTTAERKCDIYKKLGDIDLFYSLSLILPSRSD  
GPRVYIQLE"  
complement(16484. .17608)  
/gene="C33H5.12"  
complement(join(16484. .16577,16625. .16777,17217. .17324,  
17371. .17501,17555. .17608))  
/gene="C33H5.12"  
/note="similar to pre-mRNA splicing factor SRF20; coded  
for by C. elegans cDNA yk97g8.3; coded for by C. elegans  
cDNA CESK54R; coded for by C. elegans cDNA yk108h1.5;  
coded for by C. elegans cDNA yk97g8.5; coded for by C.  
elegans cDNA yk75d5.5; coded for by C. elegans cDNA  
CESK54RB; coded for by C. elegans cDNA yk108d7.5; coded  
for by C. elegans cDNA yk56g11.5"  
/codon\_start=1  
/protein\_id="AA82270.1"  
/db\_xref="GI:1086592"  
/translation="MDAKVYVGGLPSPDATSOELEETIPDFRGRIRKVVWARRPPGFAV  
EYDVRDAEDVAVLADGSRICGVARAYELSTCGRGGGGRGGGGRGGRDRSPYR  
GDRGRSRRDRGRDRGRDRGRDRGRDRGRDRGRDRGRDRGRDRGRDRGRDRGRSH  
SKSRSRSRSRSRSASPH"  
complement(18194. .18572)  
/gene="C33H5.13"  
complement(join(18194. .18314,18371. .18489,18534. .18572))  
/gene="C33H5.13"  
/note="coded for by C. elegans cDNA yk108d7.5; coded for  
by C. elegans cDNA yk56g11.5"  
/codon\_start=1  
/evidence=not experimental  
/protein\_id="AA82273.1"  
/db\_xref="GI:1086595"  
/translation="MSGAPPPPPRYNOAVYARPPGCVYVOPGVPIGGVQPAVAV  
YIQOQVIVYDCHHHHHHREHREHGGGGMCTAGAACLCFLCLSK"  
complement(18804. .20959)  
/gene="C33H5.14"  
complement(join(18804. .19064,19113. .19502,19549. .19856,  
20317. .20456,20505. .20616,20670. .20852,20896. .20959))  
/gene="C33H5.14"  
/note="similar to lymphoid cell activation antigen"  
/codon\_start=1  
/evidence=not experimental  
/protein\_id="AA82272.1"  
/db\_xref="GI:1086594"  
/translation="MSGGSSSEYVRSYVETKNNIKYVICDAGSSGTRLEFVYTLK  
PLSGGLTIDTLIHSEEPYVKKVTPGSLSPGDKPQVVEYVLPILRFAEERHYYDGLG  
ETDLLIFATGMRLLPEAQDAIILKLNQGLKSVTLRVSDSNIRIIDGAMGEGISWI

Query Match 1 1%: Score 22: DB 6: Length 4326;  
Best local Similarity 100.0%; Pred No. 1.8;  
Matches 22: Conservative 0: Mismatches 0: Indels 0: Gaps 0;  
Db 1249 ttataaaaaaatacaatttt 1270  
28452 TTATATAAAAAAATCAATTTT 28431



accession number AF007269; similar to MDJ22. GenBank  
 accession number AB006699"  
 /rpt\_type=dispersed  
 20143..23797  
 /note="function-putative transposon; transposon type-MuDR;  
 contains long flanking inverted repeats, as in Z. mays  
 MuDR, which are flanked by short direct repeats"  
 /rpt\_type=dispersed  
 20143..20146  
 /note="function-putative\_target\_duplication\_site; cognate  
 A of target duplication site flanking putative MuDR-like  
 transposon, from position 20143 to 23797"  
 /rpt\_type=flanking  
 20147..20439  
 /note="function-putative\_flanking\_inverted\_repeat; cognate  
 A of inverted repeat flanking putative MuDR-like  
 transposon, from position 20143 to 23797"  
 /rpt\_type=flanking  
 20819..23250  
 /gene="T3F12.12"  
 /note="encodes putative MuDR-A transposon protein; gene  
 model last edited on 23 Sep 97"  
 /evidence="not\_experimental  
 join(20819..21946,22024..23037,23107..23250)  
 /gene="T3F12.12"  
 /note="similar to Z. mays MuDR-A protein; functional  
 catalog ID=90"  
 /codon\_start=1  
 /evidence="not\_experimental  
 /product="putative MuDR-A-like transposon protein"  
 /protein\_id="FAB81881.1"  
 /db\_xref="GI:2565011"  
 /translation="MFDDVDVDFKTSYDGVVEYNGNSESEERIEEIVRSKKNK  
 NKKKKKDEDEPLEERFEDVDVANEDEIPIGDDIIPESDDEDPILVDRIR  
 RAMGKVFISTEFTGIEFEKALILEVLKGHNVVDKRWKKEKISFCGSGCKWVR  
 YCSYDPDOLFVWVSTGWHSCSPMGKCOLIKSPVARIPLDKILNKKPMPIIDYR  
 IKERKRVSTIPQCORGLIALKMLKKEEOPAHIRGVIEEIHSDGSAVPIIDYR  
 NKGEDVNFREYFNILITOWAGSCRPIIGDGTFLKVVGVLLAVGHDNNQIY  
 PLAAVVOSENAEMNIMFVQOLKDLNLESGSRFVILSPRSKGLISAVKELPNAHR  
 MCVRIYENLKKHAKKMDLKTIVKILAMSVNKEKYGKLNLRCEYDALYNDVNEE  
 PITRRCRYKLGSCCEDVDNNATFESNSTITKRAKSLIPMEITIRQGTIRYKKN  
 KSLRHGRTFYALKMLALEKTDADRSKVRCTHGEVYIDENGRHDIPTKQSCG  
 KWOISGICSEHSYGAMI EAGIDAEVNI SEFSDHWDSDYETATPLAGRWYLNSS  
 GIVTAPPELIPGRKKKSKKKEKFKRIKKNESKKKKKKKKEVKKLKKGIHHS  
 CEEAGHNALCKKEPEKVKPRKRSKNDNDEIEKAKKAKKNDKOTNERSKKEKS  
 KKEHMPPTITPOPSOATOE"  
 complement(23501..23793)  
 /note="function-putative\_flanking\_inverted\_repeat; cognate  
 B of inverted repeat flanking putative MuDR-like  
 transposon, from position 20143 to 23797"  
 /rpt\_type=flanking  
 23794..23797  
 /note="function-putative\_target\_duplication\_site; cognate  
 B of target duplication site flanking putative MuDR-like  
 transposon, from position 20143 to 23797"  
 /rpt\_type=flanking  
 23798..23903  
 /note="cognate A of imperfect flanking LTRs for a region  
 LTR  
 1.1%; Score 22; DB 12; Length 107183;  
 Best Local Similarity 100.0%; Pred. No. 2;  
 Matches 22; Conservative 0; Indels 0; Gaps 0;  
 QY 1195 ttgtttaaataatattttta 1216  
 Db 67091 tttgtttaaataatattttta 67070

RESULT 27  
 CEY102A5\_0  
 WPCOMMENT  
 Sequence split into 4 fragments LOCUS CEY102A5 Accession 299711  
 Fragment Name Begin End

CEY102A5\_0 1 110000  
 CEY102A5\_1 100001 210000  
 CEY102A5\_2 200001 310000  
 CEY102A5\_3 300001 389382  
 LOCUS CEY102A5 389382 bp DNA HTG 03-DEC-1998  
 DEFINITION Caenorhabditis elegans chromosome V clone Y102A5, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*, in unordered pieces.  
 299711  
 ACCESSION 299711.1 GI:3617773  
 VERSION  
 KEYWORDS HTG; PHASE1.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans.  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
 Rhabditoidea; Rhabditidae; Peloteriinae; Caenorhabditis.  
 1 (bases 1 to 389382)  
 REFERENCE  
 AUTHORS Gardner, A.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-DEC-1998) Nematode Sequencing Project, Sanger Centre,  
 Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,  
 Washington University, St. Louis, MO 63110, USA. E-mail:  
 jes@sanger.ac.uk or rw@nematode.wustl.edu  
 On Sep 17, 1998 this sequence version replaced gi:335331.  
 Order of segments is not known: 800 n/s separate segments.  
 IMPORTANT: This sequence is unfinished and does not necessarily  
 represent the correct sequence. Work on the sequence is in progress  
 and the release of this data is based on the understanding that the  
 sequence may change as work continues. The sequence may be  
 contaminated with foreign sequence from E.coli, yeast, vector,  
 phage etc.  
 \* NOTE: This is a 'working draft' sequence.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 FEATURES  
 source  
 1..389382  
 /organism="Caenorhabditis elegans"  
 /db\_xref="taxon:6239"  
 /chromosome="V"  
 /clone="Y102A5"  
 BASE COUNT 119156 a 72862 c 71081 g 122267 t 4016 others  
 ORIGIN  
 Query Match 1.1%; Score 22; DB 83; Length 110000;  
 Best Local Similarity 100.0%; Pred. No. 2;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1252 taataaaatacaatttgtt 1273  
 Db 2636 taataaaatacaatttgtt 2657

RESULT 28  
 AL139090  
 LOCUS AL139090 128468 bp DNA PRI 12-APR-2001  
 DEFINITION Human DNA sequence from clone RP11-12B13 on chromosome 6 Contains  
 STSs and GSSs, complete sequence.  
 ACCESSION AL139090  
 VERSION AL139090.11 GI:11228536  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 128468)  
 REFERENCE  
 AUTHORS Gardner, P.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 Requests: clonequests@sanger.ac.uk  
 On Nov 20, 2000 this sequence version replaced gi:11125403.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em; EMBL; Sw; SWISSPROT; Tr; TREMBL; Wp; WORMBPP; Information on the WORMBPP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Ch6>

IMPORTANT: This sequence is not the entire insert of clone RP11-12B13. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RP11-12B13 is at 1 in this sequence. The true left end of clone RP1-125N20 is at 128369 in this sequence.

The true right end of clone RP11-13018 is at 88802 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

RP11-12B13 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6.

#### FEATURES

##### Source

```

location/Qualifiers
1..128468
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone_id="RP11-12B13"
/clone_id="RPCI-11.1"
1..353
/note="match: GSS: Em:A0609536"
7..448
/note="match: GSS: Em:A0704228"
68..350
/note="match: GSS: Em:B75711"
571..1535
/note="LTR5 repeat: matches 1..966 of consensus"
1966..2087
/note="L2 repeat: matches 2574..2704 of consensus"
2290..2509
/note="L2 repeat: matches 2098..2328 of consensus"
2976..3352
/note="L2 repeat: matches 1693..2089 of consensus"
3497..3668
/note="MER20 repeat: matches 16..188 of consensus"
complement(4521..5085)
/note="match: GSS: Em:A0421033"
complement(4571..5086)
/note="match: GSS: Em:A088615"
complement(4633..5083)
/note="match: GSS: Em:A0812769"
5012..5530
/note="MER41B repeat: matches 114..601 of consensus"
5081..5634
/note="match: GSS: Em:A0544837"
6085..6174
/note="WIR repeat: matches 64..153 of consensus"
6235..6416
/note="MER5A repeat: matches 3..186 of consensus"
7146..8805
/note="LIME3 repeat: matches 4381..6140 of consensus"
8806..9399
/note="LIPAS repeat: matches 5526..6143 of consensus"
9400..9710
/note="LIME3 repeat: matches 4094..4381 of consensus"

```

```

repeat_region 9711..15846
/note="LIP6 repeat: matches 11..6143 of consensus"
15826..16269
/note="LIM4 repeat: matches 3692..4135 of consensus"
17077..17544
/note="LIM4 repeat: matches 2710..3168 of consensus"
17552..17740
/note="LIMEC repeat: matches 2408..2260 of consensus"
17741..18039
/note="AlusX repeat: matches 1..299 of consensus"
18040..18346
/note="LIMEC repeat: matches 2106..2409 of consensus"
18447..18874
/note="LIMEC repeat: matches 1469..1910 of consensus"
complement(18475..18995)
/note="match: GSS: Em:A0755181"
18938..19072
/note="LIMEC repeat: matches 1243..1377 of consensus"
19459..19514
/note="28 copies 2 mer ca 96% conserved"
20321..20675
/note="LTR16B repeat: matches 97..464 of consensus"
20952..21137
/note="AluJo repeat: matches 120..307 of consensus"
21203..21254
/note="26 copies 2 mer aa 71% conserved"
23289..23682
/note="match: GSS: Em:A081909"
23421..23472
/note="MIR repeat: matches 90..141 of consensus"
24395..24791
/note="match: GSS: Em:A027280"
24638..24848
/note="HAI1 repeat: matches 428..638 of consensus"
25008..25217
/note="105 copies 2 mer aa 55% conserved"
25328..25391
/note="32 copies 2 mer ta 70% conserved"
25482..26261
/note="LIPAS repeat: matches 5..776 of consensus"
26257..31504
/note="LIPAS repeat: matches 900..6146 of consensus"
31614..32312
/note="LIM4 repeat: matches 3918..4608 of consensus"
32398..32650
/note="LIM4 repeat: matches 4626..4909 of consensus"
32799..32963
/note="FRAM repeat: matches -2..162 of consensus"
33151..33174
/note="12 copies 2 mer tt 95% conserved"
33176..33500
/note="LTR17 repeat: matches 1..326 of consensus"
33501..33791
/note="HERV17 repeat: matches 8244..8523 of consensus"
33811..38137
/note="HERV17 repeat: matches 1651..5933 of consensus"
38135..39957
/note="HERV17 repeat: matches 1..1836 of consensus"
39958..40489
/note="LTR17 repeat: matches 257..780 of consensus"
40490..41321
/note="LIME3A repeat: matches 5261..6131 of consensus"
41354..41655
/note="AluY repeat: matches 1..302 of consensus"
41971..42279
/note="AlusX repeat: matches 1..303 of consensus"
43255..43365
/note="LIM4 repeat: matches 2169..2266 of consensus"
43366..43774
/note="MSTB repeat: matches 1..426 of consensus"
43775..44133
/note="LIM4 repeat: matches 2266..2691 of consensus"
44140..44339

```



```

misc_feature      vector_side:right"
                    6601..15485
misc_feature      /note="assembly-fragment"
                    15586..27637
misc_feature      /note="assembly-fragment"
                    27738..39958
misc_feature      /note="assembly-fragment"
                    clone_end:sp6
                    vector_side:right"
misc_feature      40059..54618
                    /note="assembly-fragment"
misc_feature      54719..82725
                    /note="assembly-fragment"
misc_feature      82826..125232
                    /note="assembly-fragment"
misc_feature      123333..167366
                    /note="assembly-fragment"
BASE COUNT      51089 a 32273 c 31977 g 51124 t 903 others
ORIGIN

```

```

Query Match      1.18; Score 22; DB 66; Length 167366;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1869 acattttattcagtaactt 1890
    |||||||||||||||||||
Db 122074 ACATTTTATTTCAGTAACCTT 122095

```

```

RESULT 30
AP002424/c
LOCUS      AP002424      171185 bp      DNA      HTG      31-MAY-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-873L22 map 18q21, WORKING
ACCESSION  AP002424
VERSION     AP002424.1 GI:8131688
KEYWORDS   HTG; HTGS-PHASE1; HTGS-DRAFT.
SOURCE     Homo sapiens DNA, clone:RP11-873L22.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 171185)
            Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
            Published Only in Database (2000) In press
            2 (bases 1 to 171185)
            Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
            Direct Submission
            Submitted (29-MAY-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            URL: http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
            ----- Genome Center
            Center: RIKEN Genomic Sciences Center(GSC)
            Center code: RIKEN
            Web site: http://hgp.gsc.riken.go.jp/
            Contact: hattori@gs.riken.go.jp
            ----- Project Information
            Center project name: HumDrafl18
            Center clone name: RP11-873L22
            ----- Summary Statistics
            Sequencing vector: PCR products; 100% of reads
            Chemistry: Dye-terminator ET-amersham; 100% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 156977 bases at least Q40
            Consensus quality: 164110 bases at least Q30
            Consensus quality: 167421 bases at least Q20
            Insert size: 169485; sum-of-ctrls

```

Quality coverage: 4.80x in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently consists of 18 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1 24923 contig of 24923 bp in length
25024 47004 contig of 21981 bp in length
47105 67070 contig of 19966 bp in length
67171 82757 contig of 15587 bp in length
82858 95544 contig of 12687 bp in length
109693 120167 contig of 10475 bp in length
120268 129255 contig of 8988 bp in length
129356 137003 contig of 7648 bp in length
137104 144376 contig of 7273 bp in length
144477 150659 contig of 6183 bp in length
150760 158896 contig of 5137 bp in length
155997 160522 contig of 4526 bp in length
160623 163263 contig of 2641 bp in length
163364 165194 contig of 1831 bp in length
165295 167704 contig of 2410 bp in length
167805 169750 contig of 1946 bp in length
169851 171185 contig of 1335 bp in length.

```

\* NOTE: This is a 'working draft' sequence. It currently consists of 18 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 24923 contig of 24923 bp in length
24924 25023: gap of 100 bp
25024 47004: contig of 21981 bp in length
47005 47104: gap of 100 bp
47105 67070: contig of 19966 bp in length
67071 67170: gap of 100 bp
67171 82757: contig of 15587 bp in length
82758 82857: gap of 100 bp
82858 95544: contig of 12687 bp in length
95545 95644: gap of 100 bp
95645 109592: contig of 13948 bp in length
109593 109692: gap of 100 bp
109693 120167: contig of 10475 bp in length
120168 120267: gap of 100 bp
120268 129255: contig of 8988 bp in length
129256 129355: gap of 100 bp
129356 137003: contig of 7648 bp in length
137004 137103: gap of 100 bp
137104 144376: contig of 7273 bp in length
144377 144476: gap of 100 bp
144477 150659: contig of 6183 bp in length
150660 150759: gap of 100 bp
150760 158896: contig of 5137 bp in length
158897 158996: gap of 100 bp
158997 160522: contig of 4526 bp in length
160523 160622: gap of 100 bp
160623 163263: contig of 2641 bp in length
163264 163363: gap of 100 bp
163364 165194: contig of 1831 bp in length
165195 165294: gap of 100 bp
165295 167704: contig of 2410 bp in length
167705 167804: gap of 100 bp
167805 169750: contig of 1946 bp in length
169751 169850: gap of 100 bp
169851 171185: contig of 1335 bp in length.
Location/Qualifiers
1..171185
/organism="Homo sapiens"

```

FEATURES  
source

```

/db_xref="taxon:9606"
/chromosome="18"
/map="18q21"
/clone="RP11-873L22"
1. 24923
misc_feature /note="assembly_fragment"
25024. .47004
misc_feature /note="assembly_fragment"
47105. .67070
misc_feature /note="assembly_fragment"
67171. .82757
misc_feature /note="assembly_fragment"
82858. .95544
misc_feature /note="assembly_fragment"
95645. .109592
misc_feature /note="assembly_fragment"
109693. .120167
misc_feature /note="assembly_fragment"
120268. .129255
misc_feature /note="assembly_fragment"
129356. .137003
misc_feature /note="assembly_fragment_clone_end:SP6_vector_side:right"
137104. .144376
misc_feature /note="assembly_fragment"
144477. .150659
misc_feature /note="assembly_fragment"
150760. .155896
misc_feature /note="assembly_fragment"
155997. .160522
misc_feature /note="assembly_fragment"
160623. .163263
misc_feature /note="assembly_fragment"
163364. .165194
misc_feature /note="assembly_fragment_clone_end:r7_vector_side:left"
165295. .167704
misc_feature /note="assembly_fragment"
167805. .169750
misc_feature /note="assembly_fragment"
169851. .171185
misc_feature /note="assembly_fragment"
BASE COUNT 52031 a 32086 c 33289 g 52078 t 1701 others
ORIGIN
Query Match 1.1% Score 22: DB 83: Length 171185;
Best Local Similarity 100.0%: Pred No. 2.1:
Matches 22: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
QY 985 tataagataatttgatatt 1006
|||||
Db 15070 TATAGTATATTTGATATT 15049
|||||
RESULT 31
AP001569/c 177097 bp DNA HTG 30-MAY-2000
LOCUS Homo sapiens chromosome 18 clone RP11-859C21 map 18q21, WORKING
DEFINITION DRAFT SEQUENCE, 32 unordered pieces.
ACCESSION AP001569
VERSION AP001569.2 GI:8117403
KEYWORDS HTG: HTGS PHASE1: HTGS DRAFT.
SOURCE Homo sapiens DNA, clone:RP11-859C21.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 177097)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens 177,097 genomic DNA of 18q21
JOURNAL Published only in database (2000) In press
REFERENCE 2 (bases 1 to 177097)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

```

```

TITLE
JOURNAL
Direct Submission
Submitted (29-MAR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsc.riken.go.jp,
URL:http://hpg.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
On May 30, 2000 this sequence version replaced gi:7380904.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hpg.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: RP11-859C21
Center clone name: RP11-859C21
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 160670 bases at least Q40
Consensus quality: 168058 bases at least Q30
Consensus quality: 171631 bases at least Q20
Insert size: 173997; sum-of-coverage
Quality coverage: 4.63x in Q20 bases; sum-of-coverage
-----
NOTE: This is a 'working draft' sequence. It currently consists of
32 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1
23893 23792 contig of 23792 bp in length
45566 21573 bp in length
58268 12602 bp in length
71274 12906 bp in length
81354 9980 bp in length
91181 9727 bp in length
97488 6207 bp in length
104902 contig of 7415 bp in length
105003 6105 bp in length
111107 contig of 7153 bp in length
111208 5644 bp in length
116952 6634 bp in length
123585 contig of 5300 bp in length
123686 5148 bp in length
128985 contig of 5148 bp in length
129086 4255 bp in length
134233 contig of 3885 bp in length
134334 3613 bp in length
138689 3667 bp in length
142573 contig of 3667 bp in length
142674 3667 bp in length
146286 contig of 3667 bp in length
150053 contig of 3667 bp in length
150154 3893 bp in length
154046 contig of 2079 bp in length
156225 contig of 2253 bp in length
156326 2253 bp in length
158578 contig of 2446 bp in length
158679 2446 bp in length
161124 contig of 2335 bp in length
161225 1879 bp in length
163559 contig of 1879 bp in length
165538 contig of 1311 bp in length
166949 contig of 1191 bp in length
167050 168240 contig of 1239 bp in length
168341 169579 contig of 1299 bp in length
169680 170978 contig of 1299 bp in length
171079 172080 contig of 1002 bp in length
172181 173514 contig of 1334 bp in length
173615 174614 contig of 1000 bp in length
174715 175976 contig of 1265 bp in length
174715 175976 contig of 1021 bp in length
176077 177097 contig of 1021 bp in length
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```

[illegible]



```

Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Strange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Stefaye,S., Tortuella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
JOURNAL Submitted (09-MAR-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 179726)
REFERENCE
AUTHORS Birten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckery,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Deatellano,K., Depayre,E., Devon,K., Dewar,K.,
Dunleat,L., Doyle,M., Ferreira,P., FitzHugh,M., Forrest,C.,
Fonke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heathford,A., Horton,L., Howland,J.C., Jones,C., Kenn,L.,
Karadas,A., Lehoczy,J., Lieu,C., Locke,K., MacDonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychalecky,J.,
Naylor,D., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Strange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Tessaye,S., Tortuella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
JOURNAL Submitted (23-MAR-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 24, 1999 this sequence version replaced gi:4432872.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES
source
1..179726
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hRPK_411_H_24"
/clone_lib="RPCK-11 human BAC library"
/map="18"
/chromosome="18"
898..1014
/rpt_family="L1MD3"
complement(1035..1333)
repeat_region
/rpt_family="Alusx"
2987..3030
/rpt_family="CT-rich"
repeat_region
/rpt_family="L1MC4"
3087..3291
4298..4330
/rpt_family="(CA)n"
repeat_region
5344..5365
/rpt_family="AT-rich"
complement(5552..5724)
repeat_region
/rpt_family="AluW"
5731..11847
/rpt_family="L1PA2"
repeat_region
11857..11891
/rpt_family="(TAA)n"
complement(11893..12002)
repeat_region
/rpt_family="AluW"
11925..11948
/note="Single-stranded coverage."
unsure
12026..12321
/rpt_family="Alusg"
complement(12605..12846)
repeat_region
/rpt_family="MIR"
13352..13560
/rpt_family="AlusC"
complement(13741..14128)
repeat_region
/rpt_family="MLT1B"
14425..14456
/rpt_family="AT-rich"
repeat_region
14457..14578
/rpt_family="MLT1F"
15496..15634
/rpt_family="THE1B"
16325..16628
/rpt_family="AluYD8"
16753..17086
/rpt_family="MER7A"
17543..17601
/rpt_family="MIR"
18075..18442
/rpt_family="MLT1B"
complement(18735..18882)
repeat_region
/rpt_family="MIR"
complement(18948..19083)
/rpt_family="L1MB8"
complement(19190..19426)
/rpt_family="AluDo"
19431..19750
/rpt_family="L1MB3"
20434..20555
/rpt_family="MLT1A1"
20556..20863
/rpt_family="Alusx"
20864..21080
/rpt_family="MLT1A1"
21528..21565
/rpt_family="AT-rich"
21575..21668
/rpt_family="L2"
complement(21849..21991)
/rpt_family="MER5B"
23147..23168
/rpt_family="(TTTG)n"
complement(23169..23384)
/rpt_family="AluDo"
23385..23419
/rpt_family="AT-rich"
complement(23799..23846)
/rpt_family="L2"
complement(23947..24060)
/rpt_family="L2"
24066..24206
/rpt_family="BC200"
complement(24347..24514)
/rpt_family="MER21B"
24782..25203
/rpt_family="AT-rich"
25385..26005
/rpt_family="L1MB2"
26035..26336
/rpt_family="L1MB2"
26432..26604
/rpt_family="Alusx"
26610..26645
/rpt_family="L1MB"
26645..26645
/rpt_family="AT-rich"
complement(27113..27422)
/rpt_family="Alusg"
27865..28190
/rpt_family="Alusg"
28250..28272
/rpt_family="AT-rich"
28363..28496
/rpt_family="HA1"
28797..28869
/rpt_family="Alusg"
28893..29027
/rpt_family="AT-rich"
29107..29134
/rpt_family="(TATATG)n"
29151..29182
/rpt_family="(TG)n"
complement(29821..30164)
/rpt_family="MER7A"

```

```

repeat_region complement(30236..30522)
repeat_region /rpt_family="AluB"
repeat_region /rpt_family="MIR"
repeat_region 30927..31271
repeat_region /rpt_family="MLTK"
repeat_region 32382..32437
repeat_region /rpt_family="CA"n"
repeat_region 32480..32502
repeat_region /rpt_family="(TTTG)n"
repeat_region complement(32503..32751)
repeat_region /rpt_family="AluX"
repeat_region 34312..34476
repeat_region /rpt_family="MIR"
repeat_region complement(34648..34953)
repeat_region /rpt_family="AluX"
repeat_region complement(35550..35947)
repeat_region /rpt_family="THEIB"
repeat_region 36048..36070
repeat_region /rpt_family="(TTTTA)n"
repeat_region complement(36072..36366)
repeat_region /rpt_family="AluB"
repeat_region 36525..36587
repeat_region /rpt_family="AT_rich"
repeat_region 36792..36849
repeat_region /rpt_family="AT_rich"
repeat_region complement(36903..37092)
repeat_region /rpt_family="L1MA7"
repeat_region 37502..37812

```

```

Query Match 1.1%; Score 22; DB 86; Length 179726;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 985 tataataatttgatatt 1006
|||||
Db 28319 TTTAGTATATTTGCAATTT 28298

```

```

RESULT 33
AC090408/c AC090408 182411 bp DNA HTG 16-MAR-2001
DEFINITION Homo sapiens chromosome 18 clone RP11-859C21 map 18, WORKING DRAFT
ACCESSION AC090408
VERSION AC090408.2 GI:13357356
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 182411)
Birten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Bouckgalter,B., Brown,A.,
Cammarata,J., Campolano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Deatellano,K., Dewar,K.J., Diaz,J.S.,
Dodgson,S., Fero,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heath,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatsis,A., Laroque,K., Lamazares,R., Landers,T.,
Lehoczky,J., Levine,R., Liu,G., Maclean,C., MacDonald,P.,
Marquis,N., Mathews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Punkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Riebeck,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosettli,M.,
Roy,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S., Severy,P.,
Sougnuez,C., Spencer,B., Stange-Tromann,N., Stojanovic,N.,

```

# TITLE JOURNAL COMMENT

Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,  
Travers, M., Travis, N., Trigilio, J., Vassiliou, H., Viel, R., Vo, A.,  
Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zainoun, J.,  
Zembek, L., Zimmer, A., and Zody, M.  
Direct Submission  
Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 16, 2001 this sequence version replaced gi:12958044.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/MW/RepeatMasker.html

Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
Project Information  
Center project name: 112718  
Center clone name: 859\_C21

Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 177348 bases at least Q40  
Consensus quality: 179869 bases at least Q30  
Consensus quality: 180657 bases at least Q20  
Insert size: 18111; sum-of-coverage  
Quality coverage: 6.2 in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently  
\* consists of 14 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```

1 803: contig of 803 bp in length
* 804 903: gap of 100 bp
* 904 1787: contig of 884 bp in length
* 1788 1887: gap of 100 bp
* 1888 3003: contig of 1116 bp in length
* 3004 3103: gap of 100 bp
* 3104 5498: contig of 2395 bp in length
* 5499 5598: gap of 100 bp
* 5599 7587: contig of 1989 bp in length
* 7588 7687: gap of 100 bp
* 7688 9739: contig of 2052 bp in length
* 9740 9839: gap of 100 bp
* 9840 13254: contig of 3415 bp in length
* 13255 13354: gap of 100 bp
* 13355 17738: contig of 4385 bp in length
* 17740 17839: gap of 100 bp
* 17840 25230: contig of 7391 bp in length
* 25231 25330: gap of 100 bp
* 25331 32077: contig of 6747 bp in length
* 32078 32177: gap of 100 bp
* 32178 48716: contig of 16539 bp in length
* 48717 48816: gap of 100 bp
* 48817 68998: contig of 20182 bp in length
* 68999 69098: gap of 100 bp
* 69099 103010: contig of 33912 bp in length
* 103011 103110: gap of 100 bp
* 103111 182411: contig of 79301 bp in length.
Location/Qualifiers
1..182411
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-859C21"
/clone_lib="RP11 Human Male BAC"
misc_feature 1..803

```

misc\_feature

```

misc_feature /note="assembly_fragment"
904.1787
misc_feature /note="assembly_fragment"
1888.3003
misc_feature /note="assembly_fragment"
3104.5498
misc_feature /note="assembly_fragment"
5599.7587
misc_feature /note="assembly_fragment"
7688.9739
misc_feature /note="assembly_fragment"
9840.13254
misc_feature /note="assembly_fragment"
13355.17739
misc_feature /note="assembly_fragment"
17840.25230
misc_feature /note="assembly_fragment"
25331.32077
misc_feature /note="assembly_fragment"
32178.48716
misc_feature /note="assembly_fragment"
48817.68898
misc_feature /note="assembly_fragment"
69099.103010
misc_feature /note="assembly_fragment"
103111.182411
misc_feature /note="assembly_fragment"
56444 a 35140 c 35375 g 54140 t 1312 others
BASE COUNT
ORIGIN

```

```

Query Match 1.1%; Score 22; DB 78; Length 182411;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 985 tataagataattgaatt 1006
|||||
Db 154736 TATAGTATATTTCATATT 154715

```

```

RESULT 34
AC087535 196869 bp DNA HTG 26-FEB-2001
LOCUS Homo sapiens chromosome 18 clone RP11-86912 map 18, WORKING DRAFT
DEFINITION AC087535.2 GI:13123259
VERSION AC087535.2
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 196869)
REFERENCE 1
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 18, clone RP11-86912
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 196869)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A.,
Camarata,J., Campopiano,A., Choquet,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Fero,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., Lacroque,K., Lamazares,R., Landers,T.,
Lehoczky,J., Levine,R., Liu,G., MacLean,C., MacDonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrum,J., Menes,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Riedack,M., Riley,R., Rise,C., Rogov,P., Roman,J., Roselli,M.,
Roy,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S., Severy,P.,

```

# TITLE JOURNAL COMMENT

Sounguez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilsson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Submitted (06-JAN-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Feb 25, 2001 this sequence version replaced g1:12043614.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MITB  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
Project Information  
Center project name: L12339  
Center clone name: 869\_L\_2  
Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 189512 bases at least Q40  
Consensus quality: 193548 bases at least Q30  
Consensus quality: 194822 bases at least Q20  
Insert size: 198000; agarose-fp  
Insert size: 195669; sum-of-contigs  
Quality coverage: 5.4 in Q20 bases; agarose-fp  
Quality coverage: 5.5 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 13 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 11489: contig of 11489 bp in length
11490 11589: gap of 100 bp
11590 12389: contig of 800 bp in length
12390 12489: gap of 100 bp
12490 15301: contig of 2812 bp in length
15302 15401: gap of 100 bp
15402 17297: contig of 1896 bp in length
17298 17397: gap of 100 bp
17398 22220: contig of 4823 bp in length
22221 22320: gap of 100 bp
22321 27991: contig of 5671 bp in length
27992 28091: gap of 100 bp
28092 31242: contig of 3151 bp in length
31243 31342: gap of 100 bp
31343 41247: contig of 9905 bp in length
41248 41347: gap of 100 bp
41348 49982: contig of 8635 bp in length
49983 50082: gap of 100 bp
50083 64395: contig of 14313 bp in length
64396 64495: gap of 100 bp
64496 79369: contig of 14874 bp in length
79370 79469: gap of 100 bp
79470 151119: contig of 71650 bp in length
151120 151219: gap of 100 bp
151220 196869: contig of 45650 bp in length.

```

## FEATURES

### source

```

1. 196869
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-86912"
/clone_lib="RP11-86912 Human Male BAC"

```

```
misc_feature 1..11489 /note="assembly-fragment
clone_end:5p6
vector_side:left"
misc_feature 11590..12389 /note="assembly-fragment"
misc_feature 12450..15301 /note="assembly-fragment"
misc_feature 15402..17297 /note="assembly-fragment"
misc_feature 17398..22220 /note="assembly-fragment"
misc_feature 22321..27991 /note="assembly-fragment"
misc_feature 28092..31242 /note="assembly-fragment"
misc_feature 31343..41247 /note="assembly-fragment"
misc_feature 41348..49982 /note="assembly-fragment"
misc_feature 50083..64395 /note="assembly-fragment"
misc_feature 64496..79369 /note="assembly-fragment"
misc_feature 79470..151119 /note="assembly-fragment"
misc_feature 151200..196869 /note="assembly-fragment
clone_end:47
vector_side:right"
BASE COUNT 56352 a 38146 c 59166 g 61991 t 1214 others
ORIGIN
```

Query Match	1.1k	Score 22:	DB 77:	Length 196865;
Best local similarity	100.0%	Pred. No. 2.2;		
Matches 22:	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	985	tataagcataatttgatatt	1006	
Db 122196	TATTAAGTATTAATTGATATT	122217		

RESULT	35
ATCHRIV24	
LOCUS	ATCHRIV24 199362 bp DNA
DEFINITION	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 24
ACCESSION	AL161512
VERSION	AL161512.2 GI:767495
KEYWORDS	
SOURCE	thale cress, <i>Arabidopsis thaliana</i>

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Unpublished  
3 (bases 1 to 199362)  
EU Arabidopsis sequencing,project.  
Direct Submission  
Submitted (10-MAR-2000) MTPs at the Max-Planck-Institut fuer  
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:1.  
bloemendaal.bloemendaal.mpp.de,mayer@mtp5.bloemendaal.mpp.de  
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge  
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,

FEATURES	source
COMMENT	E-mail: michael.bevandbbsrc.ac.uk Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <a href="http://www.mips.biochem.mpg.de/proj/hal/">http://www.mips.biochem.mpg.de/proj/hal/</a> this fragment has an overlap with ATCHRIY23 at the 5' end and an overlap with ATCHRIY25 at the 3' end.
repeat_region	location/Qualifiers 1..199362 /organism="Arabidopsis thaliana" /variety="Columbia" /db_xref="taxon:3702" /chromosome="4" 6149..9754 /note="function=putative transposon similar to T5J8, GenBank accession number AC004044"
repeat_region	10057..10255 /note="function=unclassified similar to F18G18, GenBank accession number AC006258"
repeat_region	11363..11630 /note="function=unclassified similar to MWP19, GenBank accession number AB020753"
repeat_region	12137..13086 /note="function=unclassified similar to T26D22, GenBank accession number AF058826"
repeat_region	14059..14574 /note="function=unclassified similar to T17A5, GenBank accession number AF024504"
gene	15258..17824 /gene="AT4g08540"
exon	15258..15337 /gene="AT4g08540"
CDS	/number=1 join((15258..15337,15409..15496,15644..15737,15819..15890,16129..16253,16352..16424,16528..16589,16690..16703,17232..17322,17630..17824) /gene="AT4g08540" /note="similarity to contains EST gb:193457" /codon_start=1 /product="hypothetical protein" /protein_id="CAB77979.1" /db_xref="GI:7267496"
intron	15409..15496 /gene="AT4g08540"
intron	15497..15563 /gene="AT4g08540"
exon	15364..15737 /gene="AT4g08540"
intron	15738..15818 /gene="AT4g08540"
exon	15819..15890 /gene="AT4g08540"
intron	15891..16128 /gene="AT4g08540"
exon	16129..16253 /gene="AT4g08540"

```

intron      /number=5
            16254. .16351
            /gene="AT4g08540"
            /number=5
            16352. .16424
            /gene="AT4g08540"
            /number=6
            16425. .16527
            /gene="AT4g08540"
            /number=6
            16528. .16589
            /gene="AT4g08540"
            /number=7
            16590. .16689
            /gene="AT4g08540"
            /number=7
            16690. .16703
            /gene="AT4g08540"
            /number=8
            17074. .17231
            /gene="AT4g08540"
            /number=8
            17232. .17322
            /gene="AT4g08540"
            /number=9
            17323. .17629
            /gene="AT4g08540"
            /number=9
            17630. .17824
            /gene="AT4g08540"
            /number=10
            18266. .18349
            /note="similar to acetyl CoA carboxylase"
            /note="function=unclassified similar to T24M8, GenBank
            accession number AF077409"
            22890. .23415
            /note="function=unclassified similar to T3F12, GenBank
            accession number AC002983"
            25843. .26087
            /gene="AT4g08550"
            /note="similarity to"
            /codon_start=1
            /product="hypothetical protein"
            /protein_id="CA87980.1"
            /db_xref="GI:7267497"
            /translation="MESSPKQONTEEEKSKSEIIAKVSPKQONVEERKSAITAKE
            VSPKHNVAKEEETKRPVVEIEEEEMESIDIEEGDNNSLDEIKSVSSDD
            DSESSAETACETVYDVERSGISQESNSTSAMTEKAAIKNFVAKSEVAIVFE
            IRLSGKLSIDNAGTKDDEVSQKSGSLAMPVNIIDIPTRVLELEKISGDDVP
            PVFENKLVGSYKELVLESGELEKIKHLEETPRAPRPSPSGEDDASGADV
            ELALIVLAKPCVVKDREYKMRPNKCNFLSEAVDFLDORLEDDGPRPIVETASFL
            RIYVRAITLEATSPDKHVDYSIHSGSEFARVRIIOELHVELEDDQREKLAFFI
            NLYNNMATHSLIVGHPAGTPDRTKMPDPRYVIGCTYSISALONGITLRNORPMFN
            PMKPGVAKDRSKVALPFIAPLHFTLVCGTRSGPPLKCTTGEIDKELMRAADFLR
            CGGLRYDINARVAEISKIFDW"
            25843. .26985
            /gene="AT4g08550"
            /number=1
            26986. .27312
            /gene="AT4g08550"
            /number=1
            27313. .27440
            /gene="AT4g08550"
            /number=2
            27441. .27513
            /gene="AT4g08550"
            /number=2
            27514. .27805
            /gene="AT4g08550"

```

```

intron      /number=3
            27806. .27886
            /gene="AT4g08550"
            /number=3
            27887. .28087
            /gene="AT4g08550"
            /number=4
            31602. .31788
            /note="function=unclassified similar to K1111, GenBank
            accession number AB019223"
            31932. .32214
            /gene="AT4g08560"
            /number=1
            31932. .34681
            /gene="AT4g08560"
            /note="similarity to"
            /codon_start=1
            /product="putative self-incompatibility RNA-binding
            protein"
            /protein_id="CA87981.1"
            /db_xref="GI:7267498"
            /translation="MKNLSIFMFVFSICMGVSRARIRIANELKFKNLMKRCYSKD
            DVLGPHILPIGHHFDYGTGNTFMCTLNKGPNYIHQSTAPKLEIYVGVFS

```

```

Query Match      1.18; Score 22; DB 13; Length 199362;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1195 ttgttcaaatatcttta 1216
Db 80089 TTTGTTAAATATATCTTTA 80110
|||||

```

```

RESULT 36
AP001592
LOCUS      AP001592          DNA          29-Jul-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-869L2 map 18q21, WORKING
DRAFT SEQUENCE, 11 unordered pieces.
ACCESSION AP001592.3  GI:9588073
VERSION    HTE; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS   Homo sapiens DNA, Clone:RP11-869L2.
SOURCE      Homo sapiens
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 200774)
            Hattoni M., Ishii K., Toyoda A., Taylor T.D., Hong-Seog P.,
            Fujiyama A., Yada T., Totoki Y., Watanabe H. and Sakaki Y.
            Homo sapiens 200,774 genomic DNA of 18q21
            Published Only in database (2000) In press
            2 (bases 1 to 200774)
            Hattoni M., Ishii K., Toyoda A., Taylor T.D., Hong-Seog P.,
            Fujiyama A., Yada T., Totoki Y., Watanabe H. and Sakaki Y.
            Direct Submission
            Submitted (31-MAR-2000) Masahiro Hattoni, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
            (E-mail: hattoni@gsc.riken.go.jp, URL: http://hnp.gsc.riken.go.jp/,
            Tel: 81-42-778-9923, Fax: 81-42-778-9924)
            On Jul 31, 2000 this sequence version replaced g1:8117416.
            ----- Genome Center
            Center: RIKEN Genomic Sciences Center (GSC)
            Center code: RIKEN
            Web site: http://hnp.gsc.riken.go.jp/
            Contact: hattoni@gsc.riken.go.jp
            ----- Project Information
            Center project name: HumPrat18
            Center clone name: RP11-869L2
            ----- Summary Statistics
            Sequencing vector: PCR products; 100% of reads

```

Chemistry: Dye-terminator ET-amersham: 100% of reads  
Assembly program: Phrap: version 0.990329

Consensus quality: 197367 bases at least Q40

Consensus quality: 198735 bases at least Q30

Consensus quality: 199244 bases at least Q30

Insert size: 199774; sum-of-ctrls

Quality coverage: 9.59x in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently consists of 11 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 36889 contig of 36889 bp in length
36990 73347 contig of 36358 bp in length
73448 98904 contig of 25457 bp in length
99005 124595 contig of 25591 bp in length
124696 145130 contig of 20435 bp in length
145231 163732 contig of 18502 bp in length
163833 181811 contig of 17979 bp in length
181812 181911 contig of 17979 bp in length
181912 190895 contig of 8984 bp in length
190896 190995 contig of 100 bp
190996 198298 contig of 7303 bp in length
198299 198398 contig of 100 bp
198399 199648 contig of 1250 bp in length
199649 199748 contig of 100 bp
199749 200774 contig of 1026 bp in length.
Location/Qualifiers
1.200774
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q21"
/clone="RP11-869L2"
1.36889
/note="assembly-fragment clone_end:T7 vector_side:left"
36990.73347
/note="assembly-fragment"
73448.98904
/note="assembly-fragment"
99005.124595
/note="assembly-fragment"
124696.145130
/note="assembly-fragment"
145231.163732

```

# FEATURES

## SOURCE

```

1.200774
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q21"
/clone="RP11-869L2"
1.36889
/note="assembly-fragment clone_end:T7 vector_side:left"
36990.73347
/note="assembly-fragment"
73448.98904
/note="assembly-fragment"
99005.124595
/note="assembly-fragment"
124696.145130
/note="assembly-fragment"
145231.163732

```

```

misc_feature /note="assembly-fragment"
163833.181811
misc_feature /note="assembly-fragment"
181912.190895
misc_feature /note="assembly-fragment"
190996.198298
misc_feature /note="assembly-fragment"
198399.199648
misc_feature /note="assembly-fragment clone_end:SP6 vector_side:left"
199749.200774
BASE COUNT 61549 a 39812 c 38979 g 59434 t 1000 others
ORIGIN

```

```

Query Match 1.1% Score 22; DB 82; Length 200774;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 985 tataagataattgaattac 1006
|||||
Db 97681 TATAAGTATATATTGCAATATT 97702

```

```

RESULT 37
AC079635/c 326419 bp DNA HTG 06-SEP-2000
LOCUS Mus musculus clone Rp23-152L20, WORKING DRAFT SEQUENCE, 80
DEFINITION
unordered pieces.
ACCESSION AC079635
VERSION AC079635.1 GI:9972296
KEYWORDS HTGS, PHASE1, HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 326419)
McCombie,W.R., Baker,J.P., Bahret,A., Bal,H., Dedhia,N.N., de la
Bastide,M., Huang,E.N., King,L., Kirchoff,K.A., Miller,B.,
Nascimento,L.U., O'Shaughnessy,A.L., Preston,R.R., Rodriguez,M.A.,
Shah,R.S., Sheker,M., Spiegel,L.A., Toth,K. and Vill,M.D.
Mouse Genomic Sequence
Unpublished
2 (bases 1 to 326419)
McCombie,W.R.
Direct Submission
Submitted (06-SEP-2000) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
----- Genome Center
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
Laboratory
Center code: CSHL
Web site: http://www.cshl.org/genseq
Contact: mccombie@cshl.org
----- Project Information
Center project name: Rp23-152L20
Center clone name: Rp23-152L20

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 80 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 26314 contig of 26314 bp in length
* 26315 26386 gap of unknown length
* 26387 39292 contig of 12906 bp in length
* 39293 39364 gap of unknown length
* 39365 50228 contig of 10864 bp in length
* 50229 50300 gap of unknown length

```

*	50301	60933: contig of 1063 bp in length	*	239040	239110: gap of unknown length
*	60934	61005: gap of unknown length	*	239111	241796: contig of 2686 bp in length
*	61006	70882: contig of 9877 bp in length	*	241797	241867: gap of unknown length
*	70883	70934: gap of unknown length	*	241868	244536: contig of 2669 bp in length
*	70955	79815: contig of 8861 bp in length	*	244537	244607: gap of unknown length
*	79816	79887: gap of unknown length	*	244608	247255: contig of 2648 bp in length
*	79888	88738: contig of 8851 bp in length	*	247256	247326: gap of unknown length
*	88739	88810: gap of unknown length	*	247327	249839: contig of 2513 bp in length
*	88811	97340: contig of 8530 bp in length	*	249840	249910: gap of unknown length
*	97341	97412: gap of unknown length	*	249911	252423: contig of 2511 bp in length
*	97413	105541: contig of 8129 bp in length	*	252422	252492: gap of unknown length
*	105542	105613: gap of unknown length	*	252493	254997: contig of 2505 bp in length
*	105614	113643: contig of 8030 bp in length	*	254998	255068: gap of unknown length
*	113644	113715: gap of unknown length	*	255069	257570: contig of 2502 bp in length
*	113716	121706: contig of 7991 bp in length	*	257571	257641: gap of unknown length
*	121707	121778: gap of unknown length	*	257642	260103: contig of 2462 bp in length
*	121779	128971: contig of 7193 bp in length	*	260104	260174: gap of unknown length
*	128972	129043: gap of unknown length	*	260175	262547: contig of 2373 bp in length
*	129044	134621: contig of 5578 bp in length	*	262548	262618: gap of unknown length
*	134622	134693: gap of unknown length	*	262619	264990: contig of 2372 bp in length
*	134694	140231: contig of 5538 bp in length	*	264991	265061: gap of unknown length
*	140232	140303: gap of unknown length	*	265062	267415: contig of 2354 bp in length
*	140304	145558: contig of 5255 bp in length	*	267416	267486: gap of unknown length
*	145559	145630: gap of unknown length	*	267487	269770: contig of 2284 bp in length
*	145631	150800: contig of 5170 bp in length	*	269771	269841: gap of unknown length
*	150801	150872: gap of unknown length	*	269842	272123: contig of 2282 bp in length
*	150873	155990: contig of 5118 bp in length	*	272124	272194: gap of unknown length
*	155991	156062: gap of unknown length	*	272195	274472: contig of 2278 bp in length
*	156063	160780: contig of 4718 bp in length	*	274473	274543: gap of unknown length
*	160781	160852: gap of unknown length	*	274544	276808: contig of 2265 bp in length
*	160853	165352: contig of 4500 bp in length	*	276809	276879: gap of unknown length
*	165353	165424: gap of unknown length	*	276880	279134: contig of 2255 bp in length
*	165425	169815: contig of 4391 bp in length	*	279135	279205: gap of unknown length
*	169816	169887: gap of unknown length	*	279206	281445: contig of 2240 bp in length
*	169888	174083: contig of 4196 bp in length	*	281446	281516: gap of unknown length
*	174084	174155: gap of unknown length	*	281517	283749: contig of 2233 bp in length
*	174156	178122: contig of 3967 bp in length	*	283750	283820: gap of unknown length
*	178123	178194: gap of unknown length	*	283821	285992: contig of 2172 bp in length
*	178195	181919: contig of 3725 bp in length	*	285993	286063: gap of unknown length
*	181920	181991: gap of unknown length	*	286064	288234: contig of 2171 bp in length
*	181992	185620: contig of 3629 bp in length	*	288235	288305: gap of unknown length
*	185621	185692: gap of unknown length	*	288306	290466: contig of 2161 bp in length
*	185693	189321: contig of 3629 bp in length	*	290467	290537: gap of unknown length
*	189322	189393: gap of unknown length	*	290538	292685: contig of 2148 bp in length
*	189394	192915: contig of 3522 bp in length	*	292686	292756: gap of unknown length
*	192916	192987: gap of unknown length	*	292757	294903: contig of 2147 bp in length
*	192988	196498: contig of 3511 bp in length	*	294904	294974: gap of unknown length
*	196499	196570: gap of unknown length	*	294975	297078: contig of 2104 bp in length
*	196571	200064: contig of 3494 bp in length	*	297079	297149: gap of unknown length
*	200065	200136: gap of unknown length	*	297150	299243: contig of 2094 bp in length
*	200137	203624: contig of 3488 bp in length	*	299244	299314: gap of unknown length
*	203625	203696: gap of unknown length	*	299315	301404: contig of 2090 bp in length
*	203697	207120: contig of 3424 bp in length	*	301405	301475: gap of unknown length
*	207121	207192: gap of unknown length	*	301476	303530: contig of 2055 bp in length
*	207193	210614: contig of 3422 bp in length	*	303531	303601: gap of unknown length
*	210615	210686: gap of unknown length	*	303602	305626: contig of 2025 bp in length
*	210687	213955: contig of 3269 bp in length	*	305627	305697: gap of unknown length
*	213956	214027: gap of unknown length	*	305698	307676: contig of 1979 bp in length
*	214028	217239: contig of 3212 bp in length	*	307677	307747: gap of unknown length
*	217239	217311: gap of unknown length	*	307748	309693: contig of 1946 bp in length
*	217312	220495: contig of 3184 bp in length	*	309694	309764: gap of unknown length
*	220496	220567: gap of unknown length	*	309765	311666: contig of 1902 bp in length
*	220568	223735: contig of 3168 bp in length	*	311667	311737: gap of unknown length
*	223736	223806: gap of unknown length	*	311738	313608: contig of 1871 bp in length
*	223807	226972: contig of 3166 bp in length	*	313609	313679: gap of unknown length
*	226973	227043: gap of unknown length	*	313680	315532: contig of 1853 bp in length
*	227044	230116: contig of 3073 bp in length	*	315533	315603: gap of unknown length
*	230117	230187: gap of unknown length	*	315604	317303: contig of 1700 bp in length
*	230188	233252: contig of 3065 bp in length	*	317304	317374: gap of unknown length
*	233253	233323: gap of unknown length	*	317375	319055: contig of 1681 bp in length
*	233324	236212: contig of 2889 bp in length	*	319056	319126: gap of unknown length
*	236213	236283: gap of unknown length	*	319127	320775: contig of 1649 bp in length
*	236284	239039: contig of 2756 bp in length	*	320776	320846: gap of unknown length

\* 320847 322482: contig of 1636 bp in length  
 \* 322483 322553: gap of unknown length  
 \* 322554 324187: contig of 1634 bp in length  
 \* 324188 324258: gap of unknown length  
 \* 324259 325845: contig of 1587 bp in length  
 \* 325846 325916: gap of unknown length  
 \* 325917 326419: contig of 503 bp in length.  
 Query Match 1.1%; Score 22; DB 76; Length 326419;  
 Best Local Similarity 100.0%; Pred. No. 2.3;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 487 gaggaagaagtgtatgtatgtag 508  
 Db 252266 GAGGAAGAAGTGTATGTATGTAG 252245

RESULT 38  
 AK021970/c  
 LOCUS AK021970 2061 bp mRNA PRI 29-SEP-2000  
 DEFINITION Homo sapiens CDNA FLJ11908 f1s, clone HEMMB1000089.  
 ACCESSION AK021970 GI:10433279  
 VERSION AK021970.1  
 KEYWORDS oligo capping; f1s (full insert sequence).  
 SOURCE Homo sapiens embryo, 10 weeks whole embryo, mainly body CDNA to mRNA, clone lib:HEMB1 clone:HEMB1000089.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (stiles)  
 AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fujii,A., Hara,H., Tanase,T., Nomura,Y., Togiyasu,S., Komai,F., Hara,R., Takeuchi,K., Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahara,K., Masuno,Y. and Oshima,A.  
 COMMENT NEDO human cDNA sequencing project  
 JOURNAL Unpublished (2000)  
 REFERENCE 2 (bases 1 to 2061)  
 AUTHORS Isogai,T. and Otsuki,T.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-AUG-2000) to the DBJ/EMBL/Genbank databases. Takao Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana, Katsuzu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)  
 COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

FEATURES  
 Source location/Qualifiers  
 1..2061  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="HEMB1000089"  
 /clone\_lib="HEMB1"  
 /dev\_stage="embryo, 10 weeks"  
 /tissue\_type="whole embryo, mainly body"  
 /note="Cloning vector: pME18SFL3"

BASE COUNT 657 a 324 c 387 g 693 t  
 ORIGIN  
 Query Match 1.0%; Score 21; DB 89; Length 2061;  
 Best Local Similarity 100.0%; Pred. No. 5.1;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1261 atcaatttgccttattt 1281  
 Db 1503 ATCAATTTCCTTATT 1483

RESULT 39  
 AC006224/c  
 LOCUS AC006224 27535 bp DNA PLN 05-APR-2000  
 DEFINITION Arabidopsis thaliana chromosome II section 233 of 255 of the complete sequence. Sequence from clones MFL8.  
 ACCESSION AC006224 AE002093  
 VERSION AC006224.3 GI:6598536  
 KEYWORDS HTG.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 REFERENCE 1 (bases 1 to 27535)  
 AUTHORS Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Bentolm,M.-I., Town,C.D., Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E., Feldjany,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M., Talion,L.U., Gill,J.E., Adams,M.D., Carter,A.D., Greasy,T.H., Goodman,H.M., Somerville,C.R., Coppenhaver,G.P., Preuss,D., Meriman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and Venter,J.C.  
 COMMENT Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana  
 JOURNAL Nature 402 (6763), 761-768 (1999)  
 MEDLINE 20083487  
 PUBMED 10617197  
 REFERENCE 2 (bases 1 to 27535)  
 AUTHORS Lin,X.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA  
 COMMENT On Dec 17, 1999 this sequence version replaced gi:4531433. The sequence and annotation of chromosome 2 were merged from those of the individual clones on this chromosome 2 after removing overlaps. For detailed information, please see the TIGR web site (<http://www.tigr.org/ldb/at/at.html>).

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (<http://arthur.epm.ornl.gov/pub/graill>), GeneFinder (Phil Green, University of Washington), GenScan (Chris Burge, <http://genomic.stanford.edu/GENSCAN.html>), and NePlantGene (<http://www.cds.dtu.dk/services/NetGene2/>), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/MashU/ABI consortium for sequencing BAC clones F6P23, F5J6, T17A5, and T13J16, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Peng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: [at@igr.org](mailto:at@igr.org).  
 Location/Qualifiers



[illegible]

KEYWORDS HTG: HTGS\_PHASE2.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 43784)  
AUTHORS Adams, M. and Venter, J. C.  
TITLE Direct Submission  
JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,  
Rockville, MD, USA

COMMENT This sequence was identified as CDM:10210783 by the submitter.  
For more information on this record e-mail to fly@celera.com.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

FEATURES  
source  
1. .43784  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
BASE COUNT 13647 a 7603 c 8049 g 14485 t  
ORIGIN

Query Match 1.0%; Score 21; DB 64; Length 43784;  
Best Local Similarity 100.0%; Pred. No. 7;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1247 attataaaataatcaat 1267  
|||||  
Db 5443 ATTATATAAAATCAAT 5423

RESULT 41  
LOCUS AL161618/C  
DEFINITION Human DNA sequence from clone RP11-268F1 on chromosome 6 Contains  
ESTs, STSs, GSSs and a CpG island. Contains the MEPIA gene for  
meprin A alpha (pH4 peptide hydrolase), complete sequence.  
ACCESSION AL161618  
VERSION AL161618.17 GI:10443401  
KEYWORDS HTG; CpG island; MEPIA; meprin; peptide hydrolase.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 90130)  
AUTHORS Ramsay, H.  
TITLE Direct Submission  
JOURNAL Submitted (09-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
On Oct 1, 2000 this sequence version replaced gi:10119713.

During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone configs of human  
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping  
group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr6

IMPORTANT: This sequence is not the entire insert of clone  
RP11-268F1 it may be shorter because we sequence overlapping  
sections only end of, except for a 100 base overlap.  
The true left end of clone RP11-268F1 is at 1 in this sequence. The

true left end of clone RP3-365012 is at 90031 in this sequence.  
This sequence has been finished according to sequence map criteria  
as follows. An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated repeat sequence elements. Where the sequence is  
ambiguous, there is an annotation using the 'unsure' feature key.  
RP11-268F1 is from the library RPCT-11.1 constructed by the group  
of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBAC3.6.

FEATURES  
source  
1. .90130  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone="RP11-268F1"  
/clone\_lib="RPCT-11.1"  
404. .646  
/note="MIR repeat: matches 6. .256 of consensus"  
1603. .1672  
/note="MIR repeat: matches 76. .145 of consensus"  
complement(1904. .2544)  
/note="match: GSS: Em:AQ350405"  
2494. .2876  
/note="match: GSS: Em:AQ341945"  
3235. .3410  
/note="LIMB2 repeat: matches 5986. .6163 of consensus"  
3896. .4311  
/note="L2 repeat: matches 1198. .1641 of consensus"  
5090. .5321  
/note="L2 repeat: matches 2485. .2704 of consensus"  
6624. .6935  
/note="Char1le5 repeat: matches 2257. .2585 of consensus"  
complement(7279. .7766)  
/note="match: GSS: Em:AQ838062"  
7506. .7545  
/note="10 copies 4 mer taga 100% conserved"  
7507. .7546  
/note="20 copies 2 mer ag 77% conserved"  
7735. .7869  
/note="MIR repeat: matches 123. .261 of consensus"  
8223. .8353  
/note="MIR repeat: matches 85. .224 of consensus"  
8380. .8586  
/note="MIR repeat: matches 41. .256 of consensus"  
9125. .10036  
/note="LIP4 repeat: matches 5233. .6145 of consensus"  
10344. .10411  
/note="L2 repeat: matches 2668. .2736 of consensus"  
10563. .10936  
/note="MIF18 repeat: matches 1. .359 of consensus"  
11792. .12109  
/note="AlusX repeat: matches 1. .309 of consensus"  
12385. .12505  
/note="MIR repeat: matches 21. .141 of consensus"  
12520. .12622  
/note="L2 repeat: matches 2364. .2483 of consensus"  
13106. .13537  
/note="match: GSS: Em:B62917"  
13227. .13533  
/note="AlusX repeat: matches 1. .307 of consensus"  
14167. .16741  
/note="LIME1 repeat: matches 3476. .6154 of consensus"  
16310. .16791  
/note="match: GSS: Em:B79103"  
16904. .17074  
/note="MERS9 repeat: matches 1. .178 of consensus"  
17122. .17590  
/note="Char1le4 repeat: matches 25. .491 of consensus"  
17773. .18083  
/note="AlusX repeat: matches 1. .302 of consensus"  
18286. .18345  
/note="Alu repeat: matches 1. .48 of consensus"

```

repeat_region 18331..18382
/note="Alu repeat: matches 254..303 of consensus"
repeat_region 18383..18404
/note="L1 copies 2 mer aa 100% conserved"
repeat_region 18669..18702
/note="L7 copies 2 mer ac 82% conserved"
repeat_region 18951..19113
/note="MIR repeat: matches 2..187 of consensus"
repeat_region 19506..19603
/note="MIR repeat: matches 73..173 of consensus"
repeat_region 19689..19804
/note="MIR repeat: matches 33..147 of consensus"
misc_feature complement(20792..21297)
/note="match: GSS: Em:AQ331606"
repeat_region 20795..21041
/note="L1MB7 repeat: matches 5916..6171 of consensus"
misc_feature complement(20840..21312)
/note="match: GSS: Em:AQ837608"
complement(20948..21297)
/note="match: GSS: Em:AQ138394"
repeat_region 21082..21775
/note="L2 repeat: matches 2010..2710 of consensus"
repeat_region 21776..22863
/note="PTRS repeat: matches 768..1626 of consensus"
misc_feature 21939..22375
/note="Cpg island"
/evidence=not-experimental
repeat_region 22664..23296
/note="L1R12 repeat: matches 1..671 of consensus"
repeat_region 23327..23514
/note="MIR repeat: matches 50..247 of consensus"
repeat_region 23670..23759
/note="Charlie4 repeat: matches 1855..1945 of consensus"
repeat_region 23889..24060
/note="L2 repeat: matches 2548..2746 of consensus"
repeat_region 24065..24481
/note="MLR2G repeat: matches 1..388 of consensus"
repeat_region 24482..25021
/note="L2 repeat: matches 1967..2534 of consensus"
repeat_region 25395..25934
/note="L1ME2 repeat: matches 5590..6155 of consensus"
repeat_region 26223..26664
/note="L1PB2 repeat: matches 5278..5736 of consensus"
repeat_region 26680..27027
/note="L174 copies 2 mer at 82% conserved"
repeat_region 26685..27020
/note="L24 copies 14 mer 83% conserved"
repeat_region 26687..27022
/note="L2 copies 28 mer 83% conserved"
repeat_region 27029..27218
/note="L1PB2 repeat: matches 5784..5968 of consensus"
repeat_region 27219..27514
/note="AlusC repeat: matches 1..298 of consensus"
repeat_region 27515..27686
/note="L1PB2 repeat: matches 5968..6154 of consensus"
misc_feature complement(27834..28277)
/note="match: GSS: Em:AQ465424"
repeat_region 29379..29573
/note="65 copies 3 mer aa 59% conserved"
repeat_region 29399..29566
/note="L2 copies 14 mer 64% conserved"
repeat_region 29420..29503
/note="L2 copies 4 mer aa 64% conserved"
repeat_region 29636..29894
/note="L2 repeat: matches 1816..2093 of consensus"
repeat_region 30835..30973
/note="MIR1A2 repeat: matches 238..374 of consensus"
repeat_region 30974..31254
/note="AlusG repeat: matches 1..295 of consensus"
repeat_region 31261..31430
/note="MIR1A2 repeat: matches 1..171 of consensus"
misc_feature 33314..33623
/note="match: GSS: Em:B50816"

```

```

repeat_region 33678..34033
/note="L1ME1 repeat: matches 5791..6163 of consensus"
repeat_region 34164..34297
/note="L2 repeat: matches 2577..2710 of consensus"
repeat_region 34431..34567
/note="MIR repeat: matches 119..261 of consensus"
repeat_region 36071..36109
/note="MER5A repeat: matches 53..91 of consensus"
misc_feature 36117..36505
/note="match: GSS: Em:AQ661796"
repeat_region 36303..36489
/note="MER5A repeat: matches 1..184 of consensus"
repeat_region 36663..36837
/note="MIR repeat: matches 81..256 of consensus"
repeat_region 36911..37117
/note="MER5B repeat: matches 1..224 of consensus"
repeat_region 37915..38026

```

```

Query Match 1.0%; Score 21; DB 90; Length 90130;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 325 atggtatggtgaagagtg 345
Db 28131 ATGGTATGCTGAAGGAGTG 28111

```

```

RESULT 42
AL390960/c DNA HTG 08-APR-2001
LOCUS Homo sapiens chromosome 6 clone RP1-9B16, *** SEQUENCING IN
DEFINITION PROGRESS ***, 2 unordered pieces.
ACCESSION AL390960.15 GI:13445422
VERSION AL390960.15
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 101798)
AUTHORS Pearce/A
TITLE Direct Submission
JOURNAL Submitted (07-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT On Mar 24, 2001 this sequence version replaced gi:13396621.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: d09B16
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Consensus quality: 101489 bases at least Q40
Consensus quality: 101613 bases at least Q30
Consensus quality: 101654 bases at least Q20
Insert size: 101698; sum-of-contigs
Insert size: 110053; 10.0% error; agarose-fp
Quality coverage: 9.58x in Q20 bases; sum-of-contigs quality
coverage: 8.99x in Q20 bases; agarose-fp

```

\* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. \* This record will be updated with the finished sequence \* as soon as it is available and the accession number will

```

* be preserved.
* 1 53721: contig of 53721 bp in length
* 53722 53821: gap of 100 bp
* 53822 101798: contig of 47977 bp in length.
FEATURES
    source
        1..101798
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="6"
            /clone.lib="RP1-9B16"
            /clone.lib="RPCT-1"
            1..53721
            /note="assembly_fragment:00133"
            53822..101798
            /note="assembly_fragment:00885"
BASE COUNT    33375 a 18722 c 18303 g 31298 t    100 others
ORIGIN
Query Match      1.0%; Score 21; DB 81; Length 101798;
Best Local Similarity 100.0%; Pred No. 7.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1244 taccattataaaaaaatca 1264
    |||||||||||||||||||
Db 31123 TACATTATATAAAAAAATCA 31103

RESULT 43
AC010164 103443 bp DNA PIN 03-AUG-2000
LOCUS Arabidopsis thaliana chromosome 1 BAC F14M2 genomic sequence,
DEFINITION complete sequence.
AC010164
VERSION AC010164.2 GI:6017088
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicot; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsids.
REFERENCE
    1 (bases 1 to 103443)
        Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
        Altati,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
        Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
        Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.,
        Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
        JOURNAL Unpublished
        2 (bases 1 to 103443)
        Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
        Altati,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
        Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
        Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.S.,
        Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
        JOURNAL Unpublished
        3 (bases 1 to 103443)
        Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
        Altati,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
        Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
        Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.,
        Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
        JOURNAL Submitted (09-OCT-1999) DNA Sequencing and Technology Center,
        Stanford University, 855 California Avenue, Palo Alto, CA 94304,
        USA
        4 (bases 1 to 103443)
        Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
        Altati,H., Nguyen,M., Lam,B., Southwick,A., Ecker,J., Theologis,A.
        and Davis,R.W.
        JOURNAL Direct Submission
        Direct Submission
    
```

```

JOURNAL Submitted (03-AUG-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
On Oct 9, 1999 this sequence version replaced gi:5881525.
COMMENT
FEATURES
    source
        1..103443
            /organism="Arabidopsis thaliana"
            /db_xref="taxon:3702"
            /chromosome="1"
            /clone="F14M2"
            1..26469
            /note="Overlap with bases 53840-80309 BAC clone T3M13
            gb|AC022288."
            complement(4214..5742)
            /gene="F14M2.1"
            complement(join(4214..4560,4634..4672,4724..5005,
            5352..5742))
            /gene="F14M2.1"
            /note="Hypothetical protein"
            /codon_start=1
            /protein_id="AAE97279.1"
            /db_xref="GI:9665088"
            /translation="MMRLTDNKKEMKFIYDEKNRRRLPLELTTTTEELKLMVEDYGI
            ERSTDVELSYLSDSDCLPYVMNNROVSNFSLSCKKNTYKLCVLPNGGVDDKG
            GKEKFDLNKEPDAPSDGAYFKADKVKLEKRCRPRSKTIGLIMHANEGLSPK
            PNDLSNITRTQYGHETLYQVWEAEHTVAVNGIPEKSKYKILKTYLMLEQSSQITW
            TVERPDGYKPIRKIVYVDKRIEESVLDLSSKRTYRGTIVDQGLESSQITW
            VORHIGRTAEKSRNSRSGQPYGRILELDGVIDESMRTVGMSTNPRERPDGRKK
            KVLNFKRESAMFCDRNS"
            8770..9067
            /gene="F14M2.2"
            join(8770..8912,8998..9067)
            /gene="F14M2.2"
            /note="Unknown protein"
            /codon_start=1
            /protein_id="AAE97294.1"
            /db_xref="GI:9665103"
            /translation="MEPEVVAAGIVKRRIFKFPFGVDLALDMSTEDLVKHFSSR
            IRRRSRSGFRKPMALIKLRKAVKS"
            11673..13879
            /gene="F14M2.3"
            join(11673..11954,12050..12277,12803..12913,13736..13879)
            /gene="F14M2.3"
            /note="Hypothetical protein"
            /codon_start=1
            /protein_id="AAE97296.1"
            /db_xref="GI:9665105"
            /translation="MOQPEYERYPOGTGSPAPPPKAGVIVDPKCYCSLHPVDLAIV
            KVLKITDGNFVITNAEGNLFKRVKDPFSLHKKRIIMFGFQVTLKGLKIMTHDRY
            LVFPGSTEEVDLITVYKSNMVOITKLDVFLADNIIOKKCOVRLGCVLITFSQFY
            AGSDSDITLAQMRKKTMOSVLFKGDNFCITVNDVAFASLIVLIRLPNILLKQKN
            VILRTPEPRORBLANKRNLLIKTFSRQDIYNPTL"
            15573..19096
            /gene="F14M2.4"
            join(15573..15771,15991..16244,16350..16514,17456..17587,
            17681..17758,17853..17951,18048..18158,18267..18356,
            18443..18512,18906..19096)
            /gene="F14M2.4"
            /note="Hypothetical protein"
            /codon_start=1
            /protein_id="AAE97280.1"
            /db_xref="GI:9665089"
            /translation="MAHDEECVYNLLILGRSENGKSTGNTIIGERYEVLNFGNDM
            ORCFMFAILEDGPITINVIDTPGVDYLSKINMCLTMAEGTHAVLVLSITNRISORE
            EETFNITLOQIFDCKILDYFIVFTGDELEADNQITLDYLEEGCPFLRYLVKLOGGR
            KVFNNRTKDKGRKNOLOLAHVDTIROQNGIPTYENNNHRIKEENDKLRDEGNS
            IDAKKILAEBSAMMOKLRYVDHODTNMMIOPITMKOSTTHHERHRTIEDQITMYS
            VENLKOSAMAHEREMSQLKDTLLIKDTOTNMETAPKSNVAVHRDMCOIKDTLDL
            KDITOTINMTQKSAETVPRQMRQKDTIOQRKLTMEMKEMENKAKSELAHERGHA
            HHOSPRGRQIQONLITLDINVAFDITNNKLAKVAKERNILITFYAVLVETVNP
            KSNSTLSLSE"
            21071..21997
            /gene="F14M2.5"
    
```

CDS  
21071..21997  
/gene="F14M2.5"  
/note="Hypothetical protein"  
/codon\_start=1  
/protein\_id="AA97297.1"  
/db\_xref="GI:9665106"  
/translation="MSNIGKLTLYLGIETVIOHKDIALNONRYALKITLFEAGMIOCR  
VHTFMKGLKFAEEEDVDIATKKNVCLLYLHTRPDLISGCVLSRYKQSRV  
SHAAKOCCLRLKFTTSLGLFKRCEIKPLVIGSDSHNVDGDRSTAGHVFLLG  
ESPIITWCSKQDVALSCSAEAFMAETAKQALIMOLLESLITGYCGKVITLLDNK  
SAIALTNVPEHGRSKHRRHYFIRCVETGQIEVEHPGNEOKADILTKALGRKIF  
KMRDLVGVODVVKEDFKLGENGVISLKEKK"

gene  
22333..23736  
/gene="F14M2.6"  
/note="Hypothetical protein"  
/codon\_start=1  
/protein\_id="AA97281.1"  
/db\_xref="GI:9665090"  
/translation="MSEVAVTIEKPKREGASTSIOCLPILNTYNTVAMKMKVALK  
KTMVITETGANDEPKNDMAHLLRQALPESMVIOLELDTAKKWEALIKKNVGAER  
KARLOTLANEPRLKMDKSNIEFSGRIALASAKLNTETELKVKRKLSTP  
RKKTQIVASLEVDYDLKNTSEDFVGRITVTERVWDEEDTQDDSKIMYASSN  
RDYHSGRGRGNRYGRGRGYGRDLSQVCPKDKMGHYASQCDRLKLQETOE  
NDASTKADDELMMNEIYVLANEKHCVPKLETGDDLMWLGASNMHTGDRYFENKL  
DASVYKVRGSDSRIDIKKGTIAFTDMGTGRINDVYFIPDLKNIISLQAAES  
GCDILKGNVLTMEFDGKLLVKAESKNELRYVQGLRNRCLYLTSEASIAMHAR  
MGMNSATIKPISCTIP"

CDS  
complement(27289..29061)  
/gene="F14M2.7"  
complement(join(27289..27527,27584..27851,27941..28189,  
28293..28417,28685..28827,28913..29061))  
/gene="F14M2.7"  
/note="Hypothetical protein"  
/codon\_start=1  
/protein\_id="AA97292.1"  
/db\_xref="GI:9665101"  
/translation="MGLIRFVLLISLVIFGKTTVSQPOQAAVPCITLFGDSLV  
KNNNRRLSLARANYRPGIDIPQCTGRFRNGRTTYVDLGIETGEFMRALSTOLG  
FNNYIPYSTRIGCALIKRANFASGAGIRDEGDMGHNTSMQVBLTITAVQML  
RFRFDINELORYLSRCLFYSGMSNDLNNYTPPEYSTSYNDKTFESLNTYK  
OOLRLYOGFARKVIVYVGOIGCIPLYOLANNRNNSGTGKNEKINNAIVENTYK  
KLVDRUNGOLKAKFEVLDYSKSTYDLAVGAAYVIDPGEFVWDKCGCGNN  
GGITCLPLOTPCPDRKTYLFMDAFHPTETANILLAKSNFYSAATYYPINIOELANL"

gene  
complement(34523..35499)  
/gene="F14M2.8"  
complement(join(34523..35025,35109..35499))  
/gene="F14M2.8"  
/note="Hypothetical protein"  
/codon\_start=1  
/protein\_id="AA97282.1"  
/db\_xref="GI:9665091"  
/translation="MRTKSPSSNLKVFITGSSILLIITLARSNISSSKPKST  
NLSDEEPTKORCGCTPOCTCKMPLISDAIYHVTYNNVPTQDPEVSVSRVLD  
KSPCNLYEVLGDLSDMASLNHSGRTLPTEEQANAIATYKKNPNESEHYVDTK  
VDSOKLMEIGRSECKSVSDPRNSCDLAKDPADPVEYTDMLMDVADTGHHEA  
PGMSAITYTAGLLARNREDEGETDFVHVNRPVEDFSATFLCKYMEONGRLHET  
ISHRARAGRPCEVEVDNR"

CDS  
complement(36547..39989)  
/gene="F14M2.9"  
complement(join(36547..36792,36947..37399,37485..37910,  
38010..38483,38579..39004,39777..39989))  
/gene="F14M2.9"  
/note="Hypothetical protein"  
/codon\_start=1  
/protein\_id="AA97293.1"  
/db\_xref="GI:9665102"  
/translation="MAQLEAKGKGNQMDLLDHDNLAKIHVQGHGEGIQVKEKY  
VFEDNLKIGORLGSIGLSRKQOTFEIDPTSEYISVQGVESGSIQALKEFT  
NKTSTMGIDENGLKFSNIGKAFIIGFGEADTNLSIGVAPAPPTGEOGGS  
GALMDGDSNNNGRKVSPSLDPTETROIPLVYKSLVYKPSYNSNGRQEPVVDY  
PFTIITMGTDIVSDTSKNNRSLMRTSKGTSPTFKVAARKFVESNGALIG

## gene

100.0%  
/gene="F14M2.10"

Query Match 1.0%; Score 21; DB 12; Length 103443;  
Best Local Similarity 100.0%; Pred. No. 7.6;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1246 cattataaanaaatacaaa 1266  
Db 81108 CATTATATMAAAAAATCAAA 81128

## RESULT 44

CEY52B11A/C  
LOCUS CeY52B11A 106949 bp DNA INV 25-OCT-2000  
DEFINITION Caenorhabditis elegans cosmid Y52B11A, complete sequence.  
ACCESSION AL032654 298867  
VERSION AL032654.1 GI:3810712  
KEYWORDS HTG.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida;  
Rhabdilitida; Rhabdilitidae; Peloderiinae; Caenorhabditis.

REFERENCE  
AUTHORS none.  
TITLE Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium  
JOURNAL Science 282 (5396), 2012-2018 (1998)  
MEDLINE 99069613  
REMARK The C. elegans Sequencing Consortium.  
Erratum: [published errata appear in Science 1999 Jan  
1;283(5398):35 and 1999 Mar 26;283(5410):2103 and 1999 Sep  
3;283(5433):14931]  
2 (bases 1 to 106949)  
Lennard, N.

REFERENCE  
AUTHORS Direct Submission  
TITLE Submitted (29-OCT-1998) Nematode Sequencing Project, Sanger Centre,  
Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,  
Washington University, St. Louis, MO 63110, USA. E-mail:  
jesesanger.ac.uk or twenematode.wustl.edu

## COMMENT

Coding sequences below are predicted from computer analysis, using  
predictions from Genefinder (P. Green, U. Washington), and other  
available information.  
Current sequence finishing criteria for the C. elegans genome  
sequencing consortium are that all bases are either sequenced  
unambiguously on both strands, or on a single strand with both a  
dye primer and dye terminator reaction, from distinct subclones.  
Exceptions are indicated by an explicit note.  
IMPORTANT: This sequence is not the entire insert of clone Y52B11A.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we arrange for a small overlap between  
neighbouring submissions.  
The true left end of clone Y52B11 is at 1 in this sequence. The  
true left end of clone Y53H1 is at 99300 in this sequence. The true  
left end of clone W0289 is at 106850 in this sequence. The true  
right end of clone Y95D11 is at 89618 in this sequence. The start  
of this sequence (1..107) overlaps with the end of sequence  
AL132852.  
The end of this sequence (106850..106949) overlaps with the start  
of sequence 282064.  
For a graphical representation of this sequence and its analysis  
see: - [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=Y52B11A)  
name=Y52B11A.  
Location/Qualifiers

## FEATURES

source  
1. .106949  
/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"  
/chromosome="1"  
/clone="Y52B11A.1"  
gene  
complement(join(2282. .2336, 2386. .2574, 3299. .3372, 3797. .3853, 3997. .4077))  
/gene="Y52B11A.1"  
CDS  
complement(join(2282. .2336, 2386. .2574, 3299. .3372, 3797. .3853, 3997. .4077))  
/gene="Y52B11A.1"  
/note="predicted using GeneFinder"  
/codon\_start=1  
/protein\_id="CAA2171.1"  
/db\_xref="GI:3881157"  
/translation="MEKGFNDANYTVCCSHVLAFAIFMLIDFTLTFSGLETLIFDNPAIVYGCWMLIGFTVTRFLISMDFSEKEEPLSELNFOIYIYMTIGFLVIRLFTYSAGRYGYDDDDDKDEIKRKFLEADLDEKED"  
gene  
complement(join(4254. .4370, 5910. .6104, 6323. .6484, 6537. .6684, 7779. .7930))  
/gene="Y52B11A.2"  
CDS  
complement(join(4254. .4370, 5910. .6104, 6323. .6484, 6537. .6684, 7779. .7930))  
/gene="Y52B11A.2"  
/note="predicted using GeneFinder"  
/codon\_start=1  
/protein\_id="CAA2171.1"  
/db\_xref="GI:3881157"  
/translation="MEKGFNDANYTVCCSHVLAFAIFMLIDFTLTFSGLETLIFDNPAIVYGCWMLIGFTVTRFLISMDFSEKEEPLSELNFOIYIYMTIGFLVIRLFTYSAGRYGYDDDDDKDEIKRKFLEADLDEKED"  
gene  
complement(join(12772. .17860, 18068. .18304, 18771. .18864, 19040. .19212, 19255. .19348, 20069. .20315, 21106. .21207, 22765. .22850, 28268. .28391, 41890. .42047))  
/gene="Y52B11A.4"  
CDS  
complement(join(12772. .17860, 18068. .18304, 18771. .18864, 19040. .19212, 19255. .19348, 20069. .20315, 21106. .21207, 22765. .22850, 28268. .28391, 41890. .42047))  
/gene="Y52B11A.4"

gene  
/gene="Y52B11A.4"  
/note="cDNA EST yk33562.5 comes from this gene"  
cDNA EST yk611b4.5 comes from this gene  
cDNA EST yk64992.5 comes from this gene"  
/codon\_start=1  
/protein\_id="CAB54465.1"  
/db\_xref="GI:5824824"  
/translation="MSIHRRLTPASPDIAAGNVLYTSMPTIMRCHRPLOEYRSFDMVLAQIRDSANRRLIGDVTGADYDNRSTTVAKSOSRKHKRPVDEVRMAIDDVPERKRSSTVYKRSRLOKRIIEPRTVWEAKPGLERFLKLPDSEDTYRQFKIDEGAVVSRODSRRKRRTYKSDKSSPPVSVSTDSARSSSAPPLISDVAKMTVMEMTNASTSHKTYKLYVVDGTGKSSLSQLTISEYKNAFADEIODEYTSICIGVECDLVFEESDSCWLTNEIHAPLVYSIDSSKQAMVLEIIRPGRNLPTLVAGNKIDLERKRTYKQEVRAKAMGEGHEIETSIDVDVLLGVLAETOEAFESVLOKPSRPHRPIDEPHSAIRYSQKKKAPLNDEGGCKMSPGLFAFKWRGSSPRIETN"  
gene  
join(53250. .53435, 54132. .54237, 54282. .54492, 55105. .55261)  
/gene="Y52B11A.5"  
CDS  
join(53250. .53435, 54132. .54237, 54282. .54492, 55105. .55261)  
/gene="Y52B11A.5"  
/note="predicted using GeneFinder"  
/codon\_start=1  
/protein\_id="CAA2171.1"  
/db\_xref="GI:3881159"  
/translation="MELOFELFTLIGLVATYDYGGKGRPPAPRPPPPRGVKGATGATLMKRSOGNCIRVHQYTTPQAOCHAKGATLGLQTAERSLAKGYNLHGHKRGDAIWLGARKKSGPRAGICSPDFTFTGKRTGRTAGVSPSPDGITDGVNPGVGVSCVHGFVPPSGSTPRMPGIPHGLDDQYCEGCKNPFYFVACGRKA"  
gene  
complement(join(57484. .57732, 59510. .59659, 59717. .60043, 60839. .60963, 61010. .61208, 63235. .63417))  
/gene="Y52B11A.7"  
CDS  
complement(join(57484. .57732, 59510. .59659, 59717. .60043, 60839. .60963, 61010. .61208, 63235. .63417))  
/gene="Y52B11A.7"  
/note="predicted using GeneFinder"  
/codon\_start=1  
/protein\_id="CAB63389.1"  
/db\_xref="GI:580320"  
/translation="MTIFLIVFHVLAARSPASTHLKTLGDSNPMINTQSYLDSNHRTEQASVALMSSDYEAEPDDEIETVETNIEYEAKEASLETOKIOEPBELLDPELLNMGTETREYARILKLEAKKGLSGQVNGVDQKSRFASPPFRIAPBEKMDQIVANILSPSLSDHESADEKSSLPRLDTGQDKDSFMDLLTETGYADYDEAONAFKEDSGTLINGKEELINYSKNEFTKIGEDNGRKKKVALHVIKTEIGQGNRVGKYEKRMKDVGSPLLFTSIYASAPASOALILSPVLETSILISPAIYGSVILSPWSTYPAIFKALAOSS"  
gene  
join(70106. .70241, 70298. .70390, 70618. .70712, 70772. .70894, 71641. .71718)  
/gene="Y52B11A.8"  
/note="cDNA EST yk25747.3 comes from this gene"  
cDNA EST yk25747.3 comes from this gene  
cDNA EST yk26614.3 comes from this gene  
cDNA EST yk33367.3 comes from this gene  
cDNA EST yk45669.3 comes from this gene  
cDNA EST yk25747.5 comes from this gene  
cDNA EST yk26614.5 comes from this gene  
cDNA EST yk33367.5 comes from this gene  
cDNA EST yk45669.5 comes from this gene  
cDNA EST yk57269.3 comes from this gene  
cDNA EST yk64429.3 comes from this gene  
cDNA EST yk67110.3 comes from this gene  
cDNA EST yk67110.5 comes from this gene  
cDNA EST yk64412.5 comes from this gene  
cDNA EST yk67110.5 comes from this gene"  
/codon\_start=1  
/protein\_id="CAB63390.1"  
/db\_xref="GI:580321"

```
/db_xref="SPRMBL:Q9U256"
/translation="MRGLIVATWIFVSAASATPTTTTSPPTTTTSPQILAKLP
PVNATWEGCTDEFTKISGSEIADKRIHNSCIIODHCYCAQSGCQCFDDT
FCSLEKRRSSKSCHEDEAPLPFDLVRFGDCAVAGSPNASTTESEPAEKDDYE
Query Match      1.08; Score 21; DB 6; Length 106949;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1252 taaaaaatcaatttgt 1272
|||||
Db 13786 TAAAAAATCAATTGTT 13766

RESULT 45
AC007885 LOCUS
DEFINITION
02.G.15 map 60F-60F strain Y: cn bw sp. *** SEQUENCING IN PROGRESS
***, 65 unordered pieces.
ACCESSION
AC007885 GI:6587773
VERSION
AC007885.4
KEYWORDS
HTG; HTS_PHASE1.
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 108561)
Celinker,S.E., Agbayan,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farran,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Katta,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 108561)
Celinker,S.E., Agbayan,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farran,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Katta,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomolan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poen,L., Sequeira,A., Sethi,H., Snir,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (21-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Dec 16, 1999 this sequence version replaced q1:5670579.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (http://www.fruitfly.org/sequence/) or send email
to bugfrut@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 65 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
546: contig of 546 bp in length
547
626: gap of unknown length
627
980: contig of 354 bp in length
981
1060: gap of unknown length
1061
1965: contig of 905 bp in length
1966
2045: gap of unknown length
2046
2812: contig of 767 bp in length
```

```
2813
2893
4131: contig of 1239 bp in length
4132
4211: gap of unknown length
4212
5216: contig of 1005 bp in length
5217
5296: gap of unknown length
5297
5862: contig of 566 bp in length
5863
5943
6762: contig of 820 bp in length
6763
6843
8074: contig of 1232 bp in length
8075
8154: gap of unknown length
8155
8922: contig of 768 bp in length
8923
9002: gap of unknown length
9003
9940: contig of 938 bp in length
9941
10020: gap of unknown length
10021
10762: contig of 742 bp in length
10763
10842: gap of unknown length
10843
11655: contig of 813 bp in length
11656
11735: gap of unknown length
11736
12763: contig of 1028 bp in length
12764
12843: gap of unknown length
12844
13502: contig of 659 bp in length
13503
13582: gap of unknown length
13583
14366: contig of 784 bp in length
14367
14446: gap of unknown length
14447
15348: contig of 902 bp in length
15349
15428: gap of unknown length
15429
16186: contig of 758 bp in length
16187
16266: gap of unknown length
16267
17563: contig of 1297 bp in length
17564
17644
18806: contig of 1163 bp in length
18807
18886: gap of unknown length
18887
19660: contig of 974 bp in length
19661
19940: gap of unknown length
19941
20952: contig of 1012 bp in length
20953
21032: gap of unknown length
21033
21932: contig of 900 bp in length
21933
22012: gap of unknown length
22013
23020: contig of 1008 bp in length
23021
23100: gap of unknown length
23101
24250: contig of 1150 bp in length
24251
24330: gap of unknown length
24331
25339: contig of 1609 bp in length
25340
26019: gap of unknown length
26020
26813: contig of 794 bp in length
26814
26893: gap of unknown length
26894
28391: contig of 1498 bp in length
28392
28471: gap of unknown length
28472
30255: contig of 1784 bp in length
30256
30335: gap of unknown length
30336
31417: contig of 1062 bp in length
31418
31497: gap of unknown length
31499
32408: contig of 911 bp in length
32409
32488: gap of unknown length
32489
34656: contig of 2168 bp in length
34657
34736: gap of unknown length
34737
36436: contig of 1700 bp in length
36437
36516: gap of unknown length
36517
37871: contig of 1355 bp in length
37872
37951: gap of unknown length
37952
39392: contig of 1441 bp in length
39393
39472: gap of unknown length
39473
40775: contig of 1303 bp in length
40776
40855: gap of unknown length
40856
42282: contig of 1427 bp in length
42283
42362: gap of unknown length
42363
45714: contig of 3352 bp in length
45715
45794: gap of unknown length
45795
49260: contig of 3466 bp in length
49261
49340: gap of unknown length
49341
50796: contig of 1456 bp in length
50797
50876: gap of unknown length
```

```

* 50877 52484: contig of 1608 bp in length
* 52485 52564: gap of unknown length
* 52565 56229: contig of 3665 bp in length
* 56230 56309: gap of unknown length
* 56310 59306: contig of 2997 bp in length
* 59307 59386: gap of unknown length
* 59387 63942: contig of 4556 bp in length
* 63943 64022: gap of unknown length
* 64023 67899: contig of 3877 bp in length
* 67900 73667: gap of unknown length
* 73668 73746: gap of unknown length
* 73747 80239: contig of 6493 bp in length
* 80240 80319: gap of unknown length
* 80320 83727: contig of 3408 bp in length
* 83728 83807: gap of unknown length
* 83808 90710: contig of 6903 bp in length
* 90711 90790: gap of unknown length
* 90791 96167: contig of 5377 bp in length
* 96168 96247: gap of unknown length
* 96248 97592: contig of 1345 bp in length
* 97593 97672: gap of unknown length
* 97673 98427: contig of 755 bp in length
* 98428 98507: gap of unknown length
* 98508 99158: contig of 651 bp in length
* 99159 99238: gap of unknown length
* 99239 99970: contig of 732 bp in length
* 99971 100050: gap of unknown length
* 100051 100699: contig of 649 bp in length
* 100700 100779: gap of unknown length
* 100780 101337: contig of 758 bp in length
* 101338 101617: gap of unknown length
* 101618 102197: contig of 579 bp in length
* 102197 102276: gap of unknown length
* 102277 103083: contig of 807 bp in length
* 103084 103163: gap of unknown length
* 103164 103874: contig of 711 bp in length
* 103875 103954: gap of unknown length
* 103955 104703: contig of 749 bp in length
* 104704 104783: gap of unknown length
* 104784 105423: contig of 640 bp in length
* 105424 105503: gap of unknown length
* 105504 106188: contig of 685 bp in length
* 106189 106268: gap of unknown length
* 106269 106999: contig of 731 bp in length
* 107000 107079: gap of unknown length
* 107080 107729: contig of 650 bp in length
* 107730 107809: gap of unknown length
* 107810 108561: contig of 752 bp in length.

```

## FEATURES

```

Source
1..108561
  /organism="Drosophila melanogaster"
  /strain="Y; cn bw sp"
  /db_xref="taxon:7227"
  /chromosome="2"
  /map="60F-60P"
  /clone="BACR02G15 (D643) RCT-98 02.G.15"
  /clone_lib="RCT-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
PBACE).6"
BASE COUNT 32014 a 19549 c 19037 g 32830 t 5131 others
ORIGIN

```

```

Query Match 1.0%: Score 21; DB 60; Length 108561;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1247 attataaaaaaacaatt 1267
Db 48338 ATTATATATAAAAAATCAAAAT 48318

```

```

RESULT 46
AC005142/C
LOCUS
DEFINITION
Arabidopsis thaliana BAC T5L23 from chromosome IV, near 19 cm,
complete sequence.
ACCESSION
AC005142
VERSION
AC005142.2 GI:4263038
KEYWORDS
HIC.
SOURCE
thale cress.
  Arabidopsis thaliana
  Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
  Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
  Brassicales; Brassicaceae; Arabidopsids.
  1 (bases 1 to 116448)
  Zhong,J., Ma,P., Parnell,L.D., Chen,C.-N. and Chen,E.Y.
  BAC T5L23 from chromosome IV, position 19 cm
  2 (bases 1 to 116448)
  Zhong,J., Ma,P., Parnell,L.D., Chen,C.-N. and Chen,E.Y.
  Direct Submission
  Submitted (19-JUN-1998) Applied Biosystems Division of
  Perkin-Elmer, 850 Lincoln Centre Dr., Foster City, CA 94404, USA
  3 (bases 1 to 116448)
  Parnell,L.D. and Chen,E.Y.
  Direct Submission
  Submitted (17-FEB-1999) Applied Biosystems Division, Perkin Elmer
  Corporation, 850 Lincoln Centre Dr., Foster City, CA 94404
  Arabidopsis thaliana BAC T5L23 from chromosome IV near 19 cm
  On Feb 22, 1999 this sequence version replaced gi:3241935.
  BAC T5L23 maps to near 19 cm on the Lister & Dean RI map and is
  assigned to YAC C1C8B1. Position 1 of T5L23 is oriented toward the
  telomere and position 116473 is oriented toward the centromere. For
  more information on the mapping, sequencing and annotation of
  T5L23, please see http://www.cshl.org/Arabidopsis/T5L23-titlepage.html.
  A graphic view of our annotation is also available at this url.
  Gene models are built with exons predicted by GenScan
  (http://CCR-081.mil.edu/GENSCAN.html), MZBF
  (http://www.cshl.org/genefinder) and GRAIL
  (http://compbio.ornl.gov/tools/index.shtml) and with splice sites
  predicted by NetPhosGene
  (http://www.cbs.dtu.dk/netgene/cbsnetgene.html). Genes are
  numbered according to the scheme BAC.gene.number. Typically, these
  numbers progress from 1 upwards as one moves from position 1 of the
  BAC. Protein sequences encoded by the genes are assigned to a
  functional category with the aid of similarity searches and
  comparison to the Prosite
  (http://expasy.hcuge.ch/sprot/prosite.html) and Pfam
  (http://pfam.wustl.edu/) libraries. A description of these
  categories can be found at
  http://multilac.mips.biochem.mpg.de/Arabidopsis/. Genomic repeats are
  typically located by TBLASTX analysis and an attempt is made to
  classify the function of each repeat as either Transposon, putative
  microsatellite, LINE, direct repeat, centromeric repeat, etc.

```

## FEATURES

## Source

```

1..116448
  /organism="Arabidopsis thaliana"
  /cultivar="Columbia"
  /db_xref="taxon:3702"
  /chromosome="IV"
  /map="19 CM"
  /note="hybridizing YAC is C1C8B1"
  1..31783
  /note="overlap with F9H3, GenBank accession number
AF071527, from position 1 to 31783"
  /note="Region: overlap with F9H3"
  complement(2025..3495)
  /gene="T5L23.1"
  /note="encodes putative M-type thioredoxin; identical to

```

## misc\_feature

## gene



F9H3.15, GenBank accession number AF071527; genomic copy of EST A042639; genomic copy of EST A0630735; gene model last edited on 5 Jan 99  
 /evidence=not\_experimental  
 complement(join(2025..2341,3252..3495))  
 /gene="F5L23.1"  
 /note="similar to A. thaliana protein F21B7.7, GenBank accession number AC002560; M-type thioredoxin is also present in cyanobacteria and chloroplasts; functional catalog ID=02"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="putative M-type thioredoxin"  
 /protein\_id="A015308.1"  
 /db\_xref="GI:4263039"  
 /translation="MAAPCTSRPISLSEPTLVSSPSASLSSRRNAVLPSSG LTRLSPLASTSHQPRVSKLRAYCEAETTDIOVNDSTWISLVKATGCVV VEPMAWCGCKMIDPLVNDLQHYTKIKFKLNTDESPTPCQGVRSIPTIMLEV GGEKKDTIGAVPRTLTSLDKFLP"  
 complement(4139..4163)  
 /note="function=putative\_microsatellite"  
 /rpt\_type=tandem  
 /rpt\_unit=A  
 5120..5124  
 /note="function-terminal\_flanking\_direct\_repeat; cognate A of flanking direct repeat of LTR transposon from position 5145 to 6873; cognate B resides from position 6869 to 6873"  
 /rpt\_type=flanking  
 5120..6848  
 /note="function-LTR\_transposon"  
 /rpt\_type=dispersed  
 5123..5328  
 /note="cognate A of LTR flanking transposon from position 5145 to 6873; cognate B resides from 6665 to 6868"  
 5508..6632  
 /gene="F5L23.2"  
 /note="encodes putative reverse transcriptase; likely functions as an element of LTR transposon from position 5145 to 6873; gene model last edited on 5 Jan 99"  
 /evidence=not\_experimental  
 5508..6632  
 /gene="F5L23.2"  
 /note="likely element of LTR transposon from position 5145 to 6873"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="putative reverse transcriptase"  
 /protein\_id="A015309.1"  
 /db\_xref="GI:4263040"  
 /translation="MGRKNGKSVRRSSPDLLVYASKIHEQSGSSARAFSSIHAA HQPAYSSRHEVSDSPNSHPYLVSDHFGVLAPELLDNGSITWTIAMTTS IIAKKKLGFDVGSIPKDDDDPYCKIMRCNSMWKSLNYSKITYSLIYPTAA IKDLTYRHHKSLPRLYKLRQOIHSLRQGLDSSYHRTKOTLKEETLSLQALPRV ELLDLYRTNRYVDIFLMGINDCYDAVRSQILMKRTLPSEVPNNIIDDELSARIS TTPGATSSVFAVSNOSQSYLNGDPYCKKRPVCTGSRGHEVDTCKKHGIPYTSK SKQINTELOIVDPFTKLPSHFQRLISKMGILNIPVS"  
 6640..6843  
 /note="cognate B of LTR flanking transposon from position 5145 to 6873; cognate A resides from 5150 to 5353"  
 6844..6848  
 /note="function-terminal\_flanking\_direct\_repeat; cognate B of flanking direct repeat of LTR transposon from position 5145 to 6873; cognate A resides from position 5145 to 5149"  
 /rpt\_type=flanking  
 complement(7885..7904)  
 /note="function=putative\_microsatellite"  
 /rpt\_type=tandem  
 /rpt\_unit=AT  
 complement(7905..8818)  
 /note="function=unclassified; similar to MX110, GenBank accession number AB005248"

gene  
 9482..10923  
 /gene="F5L23.3"  
 /note="encodes hypothetical protein; identical to F9H3.17, GenBank accession number AF071527; gene model last edited on 5 Jan 99"  
 /evidence=not\_experimental  
 join(9482..9639,9745..9847,10690..10923)  
 /gene="F5L23.3"  
 /note="identical to F9H3.17, GenBank accession number AF071527; similar to F21B7.5, GenBank accession number AC002560; functional catalog ID=99"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="hypothetical protein"  
 /protein\_id="A015310.1"  
 /db\_xref="GI:4263041"  
 /translation="WKLKTRIGGLVRLAFCALALAIYVITSRRASFLAISEA KTYDMAAKFYVIANAVSVSFVLELPKESLMEFVVVLDLVMTLSTLSAFLA VAOVKRKNMAGMLPFCGVPRKPDITGALVAVLYVLLLYSLHAAVDFPL LQKS"  
 12953..18635  
 /gene="F5L23.4"  
 /note="encodes putative glucan synthase component; identical to F9H3.18, GenBank accession number AF071527; gene model last edited on 5 Jan 99"  
 /evidence=not\_experimental  
 join(12953..14823,14984..17006,17187..18635)  
 /gene="F5L23.4"  
 /note="similar to 1,3-beta glucan synthase; identical to F9H3.18, GenBank accession number AF071527; functional catalog ID=01.05.99; functional catalog ID=09.01"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="putative glucan synthase component"  
 /protein\_id="A015311.1"  
 /db\_xref="GI:4263042"  
 /translation="MSLRRTVPQTPGRLAABAVGIEEPYNIIPVNNLLADPSTR PEVAAALAKTVGDLRRPYVQMRSHYDLDMALFFPGKQVNRQREHVLHA MNQMLSPPNIDISDAVVRPRKILANYSWCKSYLGKSNWISDNPSSREL LVGGLILWGEANRMPPECICIFHNASELKNTEIDPCIDBEGCOPYLPETSGEN AFLTGVAIPYDTTIOAIDESKNGVACHKRNIDINETYETDSCSKIKPDLGGS NEFKSRGKSVGKTGFVERFFFYLYRSFDRVWIALFLQAAIIVANDEKPDTSVTR QLMNALKARDVOVALTVFLTWGMRLLQAVLDAAQVLYVREKRRFFMLAKVIA AAVWIAFTVLYTINMKORODROWMSNAATTKIYQFLVAGAFVPELALALEIIPW MNPLFETNMKIFPALTMWFOGKSEVGGYLVINIKYTFVLAATFFSYFL OKVPMIKRSKILNMKLDVDYEMHOFYGDNSNFALMLPVLYLILNDIOIYATSS IYCAVYGLFDHICGIDMGQLRLRPPFASAIQPINMPREOLANRGKKEFGDGHF LKRTGEGRPFKLESNOYBAKRFALINELIAPREDIYSDREVELDELPRNSMDV TYIRMPFCLCNELLALSQARELIDAPKMLMKCKNETRCAVVBATDSTKILL SIIKVDTEHSIIYVEFOIINOSIOSEQFKTRVQLPKIYETLQKLVQVLDVEETD SCGVNVLQSLIYELATROFIEKKTTEOLDSNEGATPRDASLLFQNAIRLPASVED FVQVRLRLHLLTSRDSMSVNVLEARRIAFESSNFKMNPAPQVKKMAFVLT PYSEEVVYKSEOLRNTEDEGISLYIQTIVADENKREMHREGIKDSDSLMTTK LIDLASVYSGOTIARTEGSMYYIRALKMLAFIDSAKEMDIRGQOEGSVRNLOG ELGCGSDSFVSENDRSSISRASSSVTLXGHEYGALMKFFVVAQOITIGSKAKE

Query Match 1.0%; Score 21; DB 12; Length 116448;

Best Local Similarity 100.0%; Pred.No. 7.7;

Matches 21; Mismatches 0; Indels 0; Gaps 0;

Oy 1127 accaagaattcgaatgatt 1147

Db 82946 ACCAAGATTCGATGATT 82926

RESULT 47  
 AF289077/c AF289077 145174 bp DNA 08-AUG-2000  
 LOCUS  
 DEFINITION Homo sapiens chromosome 8 clone CTD-2003M15 map 8p33-p22, complete  
 sequence.  
 ACCESSION AF289077  
 VERSION AF289077.1 GI:9739220

```

KEYWORDS      HTG.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Mammalia: Euteria; Primates: Catarrhini; Hominoidea: Homo.
AUTHORS       Taudien, S., Bleichschmidt, K., Menzel, U., Polley, A., Reichwald, K.,
              Rump, A., Schilabel, M.B., Schudy, A., Wen, G. and Rosenthal, A.
TITLE         Chromosome 8 genomic sequence
JOURNAL       Unpublished
AUTHORS       2 (bases 1 to 145174)
TITLE         Genome Sequencing Center Jena.
JOURNAL       Direct Submission
AUTHORS       Submitted (21-JUL-2000) Genome Analysis, Institute of Molecular
              Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
FEATURES       Location/Qualifiers
              1. 145174
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="8"
                /map="8p23-p22"
                /clone="CTD-2003M15"
                /size="46075"
                /standard_name="Chr8 marker"
                /note="acc. no. U92554"
                /evidence=experimental
BASE COUNT    44314 a 26862 c 28098 g 45900 t
ORIGIN
Query Match   1.0%, Score 21; DB 89; Length 145174;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1246 cattatacaaaaatcaaa 1266
Db 132182 CATTATATATAAAATCAAA 132162

RESULT 48
AC062038/c AC062038 150880 bp DNA HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 6 clone RP11-736C22, WORKING DRAFT
ACCESSION   AC062038
VERSION     AC062038.2 GI:8516764
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Euteria; Primates; Catarrhini; Hominoidea; Homo.
TITLE       The sequence of Homo sapiens clone
JOURNAL     Unpublished
AUTHORS     2 (bases 1 to 150880)
TITLE       Waterston, R.H.
JOURNAL     Direct Submission
AUTHORS     Submitted (21-APR-2000) Genome Sequencing Center, Washington
              University School of Medicine, 444 Forest Park Parkway, St. Louis,
              MO 63108, USA
COMMENT     On Jun 14, 2000 this sequence version replaced gi:7651112.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0736C22
----- Summary Statistics -----
Sequencing vector: M13; 92%
Sequencing vector: plasmid; 8%
Chemistry: Dye-primer ET; 92% of reads

```

```

Chemistry: Dye-terminator Big Dye 8% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 14128 bases at least Q40
Consensus quality: 143702 bases at least Q30
Consensus quality: 144800 bases at least Q20
Insert size: 160000; agarose-fp
Quality coverage: 4.63 in Q20 bases; agarose-fp
Quality coverage: 4.02 in Q20 bases; sum-of-contigs
----- NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1620: contig of 1620 bp in length
1621 1720: gap of unknown length
1721 3418: contig of 1698 bp in length
3419 3518: gap of unknown length
3519 5124: contig of 1606 bp in length
5125 5224: gap of unknown length
5225 7313: contig of 2089 bp in length
7314 7414: gap of unknown length
7414 9062: contig of 1649 bp in length
9063 9162: gap of unknown length
9163 11159: contig of 1997 bp in length
11160 11259: gap of unknown length
11260 13192: contig of 1933 bp in length
13193 13292: gap of unknown length
13293 15663: contig of 2071 bp in length
15664 17082: gap of unknown length
17083 17182: contig of 1619 bp in length
17183 21212: gap of unknown length
21213 21312: gap of unknown length
21313 24147: contig of 2835 bp in length
24148 24247: gap of unknown length
24249 27695: contig of 3448 bp in length
27696 27795: gap of unknown length
27796 31536: contig of 3741 bp in length
31537 31636: gap of unknown length
31637 36118: contig of 4482 bp in length
36119 36218: gap of unknown length
36219 44091: contig of 7873 bp in length
44092 44191: gap of unknown length
44192 53451: contig of 9260 bp in length
53452 53552: gap of unknown length
53553 61968: contig of 8417 bp in length
61969 62068: gap of unknown length
62069 72196: contig of 10128 bp in length
72197 72296: gap of unknown length
72297 82467: contig of 10171 bp in length
82468 82567: gap of unknown length
82568 94770: contig of 12203 bp in length
94771 94870: gap of unknown length
94871 105474: contig of 10604 bp in length
105475 105574: gap of unknown length
105575 116975: contig of 11401 bp in length
116976 117075: gap of unknown length
117076 132819: contig of 15744 bp in length
132820 132919: gap of unknown length
132920 150880: contig of 17961 bp in length.
Location/Qualifiers
1. 150880
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-736C22"
1. 1620
/note="assembly_name:Contig6"

```

misc\_feature

/note="assembly\_name:Contig6"

```

misc_feature 1721..3418
              /note="assembly_name:Contig8"
misc_feature 3519..5124
              /note="assembly_name:Contig9"
misc_feature 5225..7313
              /note="assembly_name:Contig10"
misc_feature 7414..9062
              /note="assembly_name:Contig12"
misc_feature 9163..11159
              /note="assembly_name:Contig13"
misc_feature 11260..13192
              /note="assembly_name:Contig14"
misc_feature 13293..15363
              /note="assembly_name:Contig15"
misc_feature 15464..17082
              /note="assembly_name:Contig16"
misc_feature 17183..21212
              /note="assembly_name:Contig17"
misc_feature 21313..24147
              /note="assembly_name:Contig18"
misc_feature 24248..27695
              /note="assembly_name:Contig19"
misc_feature 27796..31536
              /note="assembly_name:Contig20"
misc_feature 31637..36118
              /note="assembly_name:Contig21"
misc_feature 36219..44091
              /note="assembly_name:Contig22"
misc_feature 44192..53451
              /note="assembly_name:Contig23"
misc_feature 53552..61968
              /note="assembly_name:Contig24"
misc_feature 62069..72196
              /note="assembly_name:Contig25"
misc_feature 72297..82467
              /note="assembly_name:Contig26"
misc_feature 82568..94770
              /note="assembly_name:Contig27"
misc_feature 94871..105474
              /note="assembly_name:Contig28"
misc_feature 105575..116975
              /note="assembly_name:Contig29"
misc_feature 117076..132819
              /note="assembly_name:Contig30"
misc_feature 132920..150880
              /note="assembly_name:Contig31"
BASE COUNT 43476 a 31949 c 30601 g 42548 t 2306 others
ORIGIN

```

```

Query Match 1.0%; Score 21; DB 72; Length 150880;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1431 tgaagaatgttaaacatt 1451
|||||
Db 128770 TGAAGAATGTATAAACATT 128750

```

```

RESULT 49
LOCUS AL136079 151320 bp DNA HTG 25-MAR-2001
DEFINITION Homo sapiens chromosome 6 clone RP3-358E10 map q24.3-25.3, ***
SEQUENCING IN PROGRESS ***
ACCESSION AL136079
VERSION AL136079.13 GI:13443283
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEPIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 151320)
AUTHORS Lloyd,D.

```

```

TITLE Direct Submission
JOURNAL Submitted (21-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT requests: clonerequest@sanger.ac.uk
On Mar 24, 2001 this sequence version replaced gi:12580937.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: dj358E10
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M77815; 12% of reads
Sequencing vector: plasmid; 108752; 87% of reads
Chemistry: dye-terminator Big Dye; 89% of reads
Chemistry: dye-terminator Big Dye; 10% of reads
Chemistry: dye-terminator Big Dye; 0% of reads
Consensus quality: 150706 bases at least Q40
Consensus quality: 150817 bases at least Q30
Consensus quality: 150880 bases at least Q20
Insert size: 151120; sum-of-contigs
Insert size: 166819; 1.9% error; agarose-fp
Quality coverage: 10.88x in Q20 bases; sum-of-contigs quality
coverage: 10.78x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

FEATURES
Source
1..151320
1 39304: contig of 39304 bp in length
* 39305 39404: gap of 100 bp
* 39405 41914: contig of 2510 bp in length
* 41915 42014: gap of 100 bp
* 42015 151320: contig of 109306 bp in length.
Location/Qualifiers
1..151320
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="q24.3-25.3"
/clone="RP3-358E10"
/clone_lib="RPCI-3"
1..39304
/note="assembly_fragment:01530"
39405..41914
/note="assembly_fragment:02989"
42015..151320
/note="assembly_fragment:05398"

```

```

misc_feature 1..39304
/note="assembly_fragment:01530"
misc_feature 39405..41914
/note="assembly_fragment:02989"
misc_feature 42015..151320
/note="assembly_fragment:05398"
BASE COUNT 47015 a 29144 c 28365 g 46595 t 201 others
ORIGIN

```

```

Query Match 1.0%; Score 21; DB 78; Length 151320;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1244 tacattataaaaaaatca 1264
|||||
Db 120901 TACATTATATAAAAAATCA 120921

```

```

RESULT 50
LOCUS AC023306/c 153266 bp DNA HTG 07-APR-2001
DEFINITION Homo sapiens chromosome 8 clone RP11-199C5 map 8, WORKING DRAFT
SEQUENCE 5 unordered pieces.
ACCESSION AC023306

```

VERSION AC023306.5 GI:13560420  
 KEYWORDS HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 153266)  
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
 Homo sapiens chromosome 8, clone RP11-199C5  
 JOURNAL Unpublished  
 TITLE 2 (bases 1 to 153266)  
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
 Anderson, S., Baldwin, J., Barna, N., Bedalov, F., Boguslavsky, L.,  
 Boucknight, B., Brown, A., Burkett, G., Campotiano, A., Castle, A.,  
 Choquel, Y., Colangelo, M., Collins, S., Collumore, A., Cooke, P.,  
 Dearlano, K., Dewar, K., Dodge, S., Domino, M., Doyle, M.,  
 Feneator, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D.,  
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
 Grand-pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,  
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
 Klein, J., Landers, T., Largocque, K., Lehotzky, J., Levine, R.,  
 Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M.,  
 McEwan, P., McGurk, A., McKernan, K., McPheters, R., Meldrum, J.,  
 Menes, L., Milova, T., Miranda, C., Mienga, V., Morrow, J., Naylor, J.,  
 Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M.,  
 Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, J.,  
 Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,  
 Severy, P., Spencer, B., Stange-Rothman, N., Stojanovic, N.,  
 Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirelli, A.,  
 Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,  
 Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and  
 Zody, M.

TITLE Direct Submission  
 JOURNAL Submitted (11-FEB-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 COMMENT On Apr 7, 2001 this sequence version replaced gi:11693423.  
 All repeats were identified using RepeatMasker:  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 Smith, A.F.A. & Green, P. (1996-1997)

Center: Whitehead Institute/ MIT Center for Genome Research  
 Genome Center  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L5186  
 Center clone name: L5186  
 ----- Summary Statistics  
 Sequencing vector: M13; M77815; 3% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 151660 bases at least Q40  
 Consensus quality: 152433 bases at least Q30  
 Consensus quality: 152665 bases at least Q20  
 Insert size: 151000; agarose-fp  
 Insert size: 152866; sum-of-ctnigs  
 Quality coverage: 8.0 in Q20 bases; agarose-fp  
 Quality coverage: 7.9 in Q20 bas.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 5 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 17122: contig of 17122 bp in length  
 \* 17123 17223: gap of 100 bp  
 \* 17223 29433: contig of 12211 bp in length  
 \* 29434 29533: gap of 100 bp  
 \* 29534 51685: contig of 22152 bp in length  
 \* 51686 51785: gap of 100 bp

FEATURES  
 source  
 \* 51786 110606: contig of 58821 bp in length  
 \* 110607 110706: gap of 100 bp  
 \* 110707 153266: contig of 42560 bp in length.  
 Location/Qualifiers  
 1. 153266  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="8"  
 /map="8"  
 /clone="RP11-199C5"  
 /clone\_1lb="RP11-11 Human Male BAC"  
 1. 17122  
 /note="assembly-fragment"  
 /clone\_end:SP6  
 vector\_side:left"  
 17223. 29433  
 /note="assembly-fragment"  
 29534. 51685  
 /note="assembly-fragment"  
 51786. 110606  
 /note="assembly-fragment"  
 110707. 153266  
 /note="assembly-fragment"  
 clone\_end:T7  
 vector\_side:right"  
 BASE COUNT 46235 a 28641 c 28900 g 49090 t 400 others  
 ORIGIN

Query Match 1.0%; Score 21; DB 67; Length 153266;  
 Best Local Similarity 100.0%; Pred. No. 7.9;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1246 cattataaaaaaatcaaa 1266  
 ||||||||||||||||||||  
 Db 84729 CATTATAAAAAAATCAAA 84709

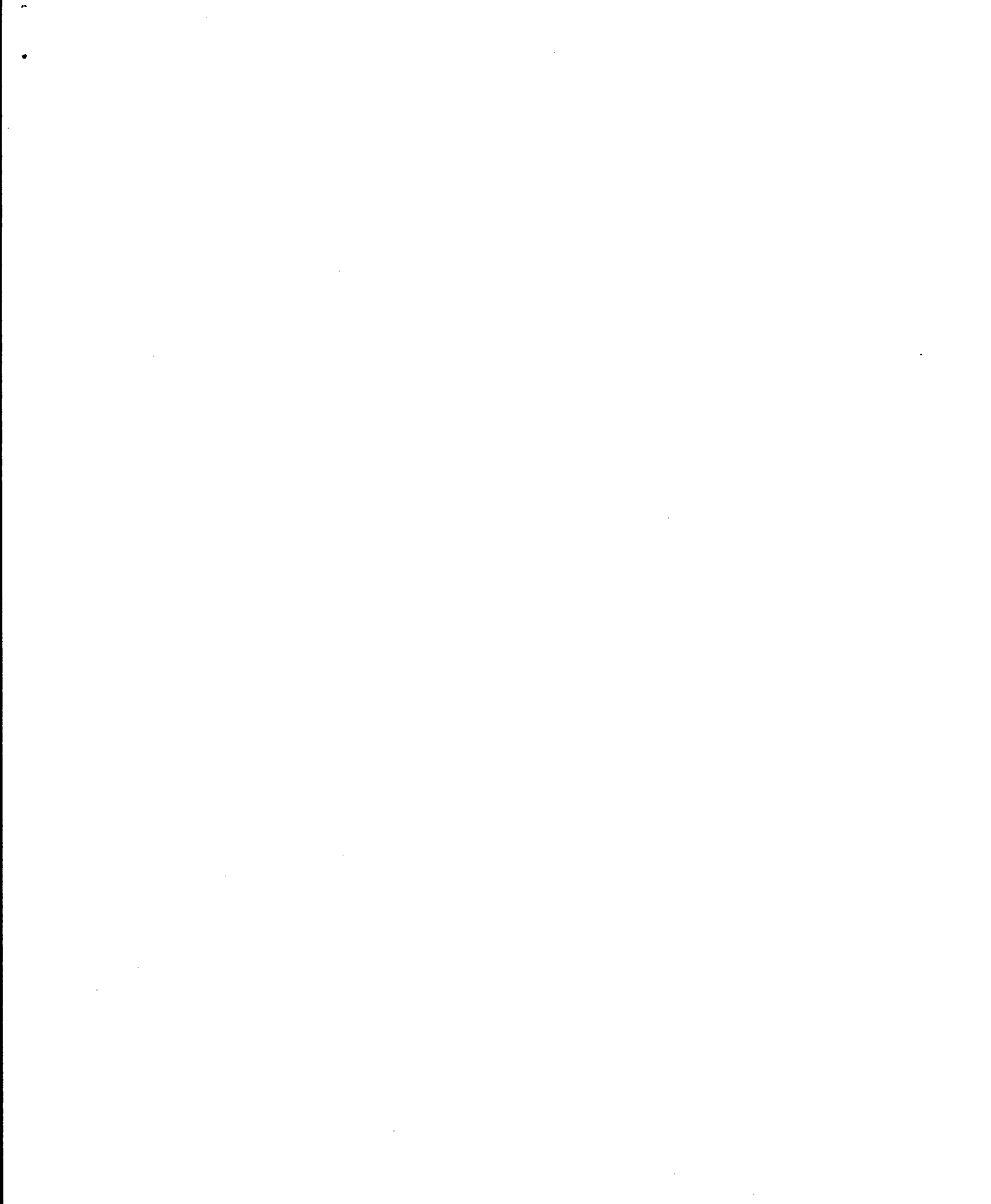
Search completed: August 19, 2001, 02:29:16  
 Job time: 22336 sec

Mon Aug 20 10:21:39 2001

us-09-284-320-56.0110.rge

---

Page 55

















C 815	15	0.7	417	21	AAC27665	Human secreted pro
C 816	15	0.7	418	21	AAC37718	Arabidopsis thalia
C 817	15	0.7	418	21	AAC21006	Human secreted pro
C 818	15	0.7	422	16	AAO79931	polycistronic DNA
C 819	15	0.7	423	18	AAE80625	Type II topoisomerase
C 820	15	0.7	426	6	AAAS0556	Sequence encoding
C 821	15	0.7	426	8	AAAT0162	Sequence of human
C 822	15	0.7	426	20	AAV88450	EST clone G174. H
C 823	15	0.7	432	6	AAAS0558	Sequence encoding
C 824	15	0.7	432	17	AAAT30259	Cotton fibre cell-
C 825	15	0.7	432	17	AAAT3042	Cotton fibre-speci
C 826	15	0.7	432	18	AAAT70049	Cotton fibre speci
C 827	15	0.7	432	18	AAE62618	Cotton fibre speci
C 828	15	0.7	432	21	AAZ35553	CDNA sequence a co
C 829	15	0.7	433	20	AAE08304	Human lung tumour
C 830	15	0.7	433	21	AAE79156	Human lung tumour
C 831	15	0.7	435	20	AAAX20102	Enterococcus faeca
C 832	15	0.7	436	18	AAV74875	Staphylococcus aur
C 833	15	0.7	436	22	AAE65287	Novel human polynu
C 834	15	0.7	437	21	AAAC27591	Human secreted pro
C 835	15	0.7	440	21	AAE10616	Fusarium venenatum
C 836	15	0.7	441	19	AAV40541	Homo sapiens secre
C 837	15	0.7	443	22	AAE97838	Human neuroblastom
C 838	15	0.7	444	10	AAAN91069	Cloned CDNA encodi
C 839	15	0.7	447	21	AAAC27920	Fusarium venenatum
C 840	15	0.7	454	21	AAE08133	Human secreted pro
C 841	15	0.7	454	21	AAAC24845	Human secreted pro
C 842	15	0.7	462	21	AAAC17849	Human secreted pro
C 843	15	0.7	467	20	AAV90149	EST clone DF218.
C 844	15	0.7	471	21	AAAC19461	Human secreted pro
C 845	15	0.7	473	20	AAAX51646	Human secreted pro
C 846	15	0.7	477	21	AAAS0279	Human LSR gene 5'
C 847	15	0.7	477	21	AAAC37179	Arabidopsis thalia
C 848	15	0.7	481	18	AAAT83867	DNA encoding a Sta
C 849	15	0.7	483	19	AAAV60043	Nucleic acid E69
C 850	15	0.7	490	21	AAAC56438	Pinus radiata tran
C 851	15	0.7	494	19	AAAV20933	Aloe capensis trnl
C 852	15	0.7	497	21	AAAA47444	Human secreted exp
C 853	15	0.7	497	21	AAZ50513	Human hybrid inter
C 854	15	0.7	497	21	AAZ50516	Human mutant hybri
C 855	15	0.7	497	21	AAZ50517	Human mutant hybri
C 856	15	0.7	500	21	AAZ50309	Hybrid interferon
C 857	15	0.7	500	21	AAZ50310	Hybrid interferon
C 858	15	0.7	500	21	AAZ50327	Hybrid interferon
C 859	15	0.7	500	21	AAZ50328	Hybrid interferon
C 860	15	0.7	500	21	AAZ50330	Mutant hybrid inte
C 861	15	0.7	500	21	AAZ50331	Mutant hybrid inte
C 862	15	0.7	500	21	AAZ50332	Mutant hybrid inte
C 863	15	0.7	500	21	AAZ50333	Mutant hybrid inte
C 864	15	0.7	500	21	AAZ50510	Human hybrid inter
C 865	15	0.7	500	21	AAZ50511	Human hybrid inter
C 866	15	0.7	500	21	AAZ50514	Human hybrid inter
C 867	15	0.7	500	21	AAZ50518	Human mutant hybri
C 868	15	0.7	500	21	AAZ50519	Human mutant hybri
C 869	15	0.7	501	8	AAAT0528	Sequence encoding
C 870	15	0.7	501	18	AAAT0528	H. pylori cell env
C 871	15	0.7	501	22	AAE93196	CDNA encoding SRT
C 872	15	0.7	503	5	AAAN0013	DNA encoding recom
C 873	15	0.7	504	8	AAAT0533	Sequence encoding
C 874	15	0.7	504	13	AAO25560	Human interferon a
C 875	15	0.7	504	13	AAO47152	Recombinant human
C 876	15	0.7	504	20	AAZ30023	DNA encoding a hum
C 877	15	0.7	504	20	AAV88369	EST clone DB518.
C 878	15	0.7	504	21	AAAB9642	Exo56 nucleotide s
C 879	15	0.7	504	21	AAE93034	PMON20442 encoding
C 880	15	0.7	507	20	AAZ30021	DNA encoding a hum
C 881	15	0.7	507	20	AAZ30022	DNA encoding a hum
C 882	15	0.7	507	20	AAZ30022	Nucleic acid encod
C 883	15	0.7	507	20	AAZ30022	Nucleic acid encod
C 884	15	0.7	507	21	AAZ30022	Human cancer assoc
C 885	15	0.7	507	21	AAZ30022	PMON36422 encoding
C 886	15	0.7	507	21	AAZ30022	PMON36422 encoding
C 887	15	0.7	507	21	AAZ30022	PMON36422 encoding

CDNA encoding SRT  
Human secreted exp  
Human lung tumour  
Zea mays DNA fragm  
Arabidopsis thalia  
Extended human sec  
Nucleotide sequenc  
Human secreted pro  
Human interferon a  
Cat flea hindgut a  
Modified human gra  
Interferon alpha a  
Arabidopsis thalia  
Secreted protein E  
Arabidopsis thalia  
Zea mays DNA fragm  
Arabidopsis thalia  
Arabidopsis thalia  
Human secreted pro  
Cat flea hindgut a  
Cat flea hindgut a  
DNA encoding a Sta  
DNA marker B. Lac  
EST clone FB264  
Arabidopsis thalia  
Human secreted exp  
Human pancreatic c  
Polynucleotide seq  
Human interferon w  
Coding sequence fo  
Exo82 nucleotide s  
Arabidopsis thalia  
Sequence encoding  
Arabidopsis thalia  
ACNPV ORF 68, res1  
Human colon cancer  
Human secreted pro  
DNA encoding inter  
Fusarium venenatum  
CDNA encoding aspa  
Arabidopsis thalia  
C. symbiosum open  
Human secreted pro  
C. symbiosum open  
Human colon cancer  
Human secreted pro  
N. meningitidis pa  
Arabidopsis thalia  
Human secreted pro  
Lobliolly pine SSR  
Human prostate can  
Arabidopsis thalia  
Arabidopsis thalia  
Mature native gran  
Nucleotide sequenc  
Polynucleotide seq  
B. burgdorferi ant  
Translatational fusi  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Human colon cancer  
Human IFNalpha 2C  
pMD19B-lfN HindIII  
H. pylori putative  
Natural human inte  
Human gene expres  
HIV-1 group O env  
Recombinant p60-8p  
HIV-1 group O env  
ET-related gene fr  
ET-related gene fr



QY 661 gtgtacatgatattcaagctgtcgtctcgtcaataagcatctagccaagatcatct 720  
 |||||||  
 Db 661 gtgtacatgatattcaagctgtcgtctcgtcaataagcatctagccaagatcatct 720  
 QY 721 cctgattatattcaacgtgagacgtgcagatgttgatgaattggagagcgataggggaa 780  
 |||||||  
 Db 721 cctgattatattcaacgtgagacgtgcagatgttgatgaattggagagcgataggggaa 780  
 QY 781 gactcgaacaattcagaagatgcttcaagatcctgtgtaacgtctgcgaagaattgca 840  
 |||||||  
 Db 781 gactcgaacaattcagaagatgcttcaagatcctgtgtaacgtctgcgaagaattgca 840  
 QY 841 gatgaatgacagtccttattgtgtggaatgcagtggtgagatgaatcctgtcgaatgca 900  
 |||||||  
 Db 841 gatgaatgacagtccttattgtgtggaatgcagtggtgagatgaatcctgtcgaatgca 900  
 QY 901 tttagaacctccctcattaggaagacaagagactacccctgagcaaaacaagcgaagac 960  
 |||||||  
 Db 901 tttagaacctccctcattaggaagacaagagactacccctgagcaaaacaagcgaagac 960  
 QY 961 ccagcaagtcctcctaataacctgtgcatataagtaatttgatataatccgtggtttcaac 1020  
 |||||||  
 Db 961 ccagcaagtcctcctaataacctgtgcatataagtaatttgatataatccgtggtttcaac 1020  
 QY 1021 atgtgacttgatataatgatcgtccttgctgctgtgcatatcacaccttacaataat 1080  
 |||||||  
 Db 1021 atgtgacttgatataatgatcgtccttgctgctgtgcatatcacaccttacaataat 1080  
 QY 1081 tggaaacttgatcccgatagatagatcatatttaagaagaaacccaagaagtttga 1140  
 |||||||  
 Db 1081 tggaaacttgatcccgatagatagatcatatttaagaagaaacccaagaagtttga 1140  
 QY 1141 atgattgaatgttaccctgtgcccagaattagaagaaggggttgaagaattgagctgttct 1200  
 |||||||  
 Db 1141 atgattgaatgttaccctgtgcccagaattagaagaaggggttgaagaattgagctgttct 1200  
 QY 1201 taaatatactttagtgtgtcttaagaagatagatgacttacttactatataaaaaaa 1260  
 |||||||  
 Db 1201 taaatatactttagtgtgtctttaaagaagatagatgacttacttactatataaaaaaa 1260  
 QY 1261 atcaaatctgtcttatttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1320  
 |||||||  
 Db 1261 atcaaatctgtcttatttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1320  
 QY 1321 attgacgtgaatcccaactgtgtgtatagatccatattgcttgaattatgatatagc 1380  
 |||||||  
 Db 1321 attgacgtgaatcccaactgtgtgtatagatccatattgcttgaattatgatatagc 1380  
 QY 1381 catttaatacattgattcattctgtttaaattgaatttgaataatgcaactgagaagaat 1440  
 |||||||  
 Db 1381 catttaatacattgattcattctgtttaaattgaatttgaataatgcaactgagaagaat 1440  
 QY 1441 gtaaaacatttaagatgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1500  
 |||||||  
 Db 1441 gtaaaacatttaagatgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1500  
 QY 1501 ttacgaatgttacttcttaccacagatagggtgaataatcatttgggtctatgttata 1560  
 |||||||  
 Db 1501 ttacgaatgttacttcttaccacagatagggtgaataatcatttgggtctatgttata 1560  
 QY 1561 ctatgaacaatttgtaaatgtcttaatttgaatttaataactcgtgaacaagaagaagaag 1620  
 |||||||  
 Db 1561 ctatgaacaatttgtaaatgtcttaatttgaatttgaataactcgtgaacaagaagaagaag 1620  
 QY 1621 ttttaacttagtagtagccctaaatattgattgtctatataatcgtcttagtttggaa 1680  
 |||||||  
 Db 1621 ttttaacttagtagtagccctaaatattgattgtctatataatcgtcttagtttggaa 1680  
 QY 1681 cgtgtatcgaataacagagagacagctgtttttaaccccttcgtcaagtttgttgaact 1740  
 |||||||  
 Db 1681 cgtgtatcgaataacagagagacagctgtttttaaccccttcgtcaagtttgttgaact 1740

QY 1741 acatgggctaataatgataactaaataactacattgatctcaagaagaactagctgtgt 1800  
 |||||||  
 Db 1741 acatgggctaataatgataactaaataactacattgatctcaagaagaactagctgtgt 1800  
 QY 1801 gagtataatgagcttcttcatlatacacaaataatccctgaagagacatttggagcat 1860  
 |||||||  
 Db 1801 gagtataatgagcttcttcatlatacacaaataatccctgaagagacatttggagcat 1860  
 QY 1861 aatataaacaatttatttcatgtaacatttcccccgtgtgaagtactacatggtttgtg 1920  
 |||||||  
 Db 1861 aatataaacaatttatttcatgtaacatttcccccgtgtgaagtactacatggtttgtg 1920  
 QY 1921 tacaacttactctataagaataatgaatgagtggtggtgaattcacttattatgttga 1980  
 |||||||  
 Db 1921 tacaacttactctataagaataatgaatgagtggtggtgaattcacttattatgttga 1980  
 QY 1981 gtggaaccaatgtctatcaagagtgacaaataaagttaattgattcccaaac 2033  
 |||||||  
 Db 1981 gtggaaccaatgtctatcaagagtgacaaataaagttaattgattcccaaac 2033  
 |||||||

RESULT 2  
 AAF97893  
 ID AAF97893 standard; CDNA; 2108 BP.  
 AC AAF97893:  
 XX 01-JUN-2001 (first entry)  
 D7  
 DE Human secreted protein cDNA, SEQ ID NO: 20.  
 XX  
 KW Human; secreted protein; immunomodulatory; antisclerotic;  
 KW dermatological; antiinflammatory; anti-HIV; cytostatic; cardiant;  
 KW vascular; anti-angiogenic; ophthalmological; neuroprotectant;  
 KW neotropic; anticonvulsant; antialzheimers; antiparkinsonian;  
 KW antimicrobial; vulnereary; vaccine; gene therapy; cancer;  
 KW protein coordinate data; infection; ss.  
 KW  
 XX Homo sapiens.  
 OS  
 XX  
 PN WO200121658-A1.  
 XX  
 PD 29-MAR-2001.  
 XX  
 PF 22-SEP-2000; 2000MO-US26013.  
 XX  
 PR 24-SEP-1999; 99US-0155709.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Ni J, Baker KP, Birse CE, Ebner R, Fiscella M, Komatsoulis GA;  
 PI Lafleur DM, Moore PA, Olsen HS, Rosen CA, Ruben SA, Soppet DR;  
 PI Young PE, Wei P, Florence KA;  
 XX  
 DR WPI: 2001-235311/24.  
 XX  
 PT Nucleic acids encoding 32 human secreted polypeptides, useful for  
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's  
 PT disease and diabetic retinopathy.  
 PT  
 PS Claim 1: Page 726-727; 890pp; English.  
 XX  
 CC The present sequence encodes one of 32 novel human secreted polypeptides.  
 CC The nucleic acid molecules and polypeptides they encode may be used in  
 CC the prevention, diagnosis and treatment of diseases such as  
 CC immune disorders (e.g. multiple sclerosis, systemic lupus erythematosus  
 CC and human immuno-deficiency virus (HIV) infections), hyperproliferative  
 CC disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases  
 CC (e.g. Schmitz syndrome, Chaga's cardiomyopathy and coronary  
 CC arteriosclerosis), angiogenic disorders (e.g. corneal graft  
 CC neovascularisation and diabetic retinopathy), neurological disorders  
 CC (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),  
 CC infectious diseases and/or for promoting wound healing, regeneration







[illegible]

Pt	EP1067182-A2.
Pd	10-JAN-2001.
Pf	07-JUL-2000; 2000EP-0114090.
Px	08-JUL-1999; 99JP-01941179.
Px	11-MAY-2000; 2000JP-0118775.
Pr	02-JAN-2000; 2000JP-0183766.
Pa	(HELI-) HELIX RES INST.
Pi	Oka T, Isegai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
Dk	WPI; 2001-093989/11.
P-PSDB	AAB88347.
Pt	Nucleic acids encoding secretory proteins/membrane proteins, useful in gene therapy or as candidate target molecules in drug development -
Px	Claim 1; SEQ ID 61; 609pp + CD ROM; English.
Px	This invention relates to nucleic acid sequences AAF93744 - AAF93916 which encode human secretory or membrane proteins represented by AAB88347 - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the invention. The invention also includes methods for the production of antibodies directed against the proteins, and cDNA sequences, which can be used in vaccines. The polynucleotide sequences can be used in gene therapy. The polynucleotide sequences and the proteins they encode may be used in the prevention, treatment and diagnosis of diseases associated with inappropriate secretory protein/membrane protein expression. The nucleic acids and complementary sequences may also be used as DNA probes in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect and quantitate the presence of similar nucleic acid sequences in samples. They may also be used to study the expression and function of secretory proteins/membrane polypeptides and their role in metabolism. The polypeptides may be used as antigens in the production of antibodies against them and in assays to identify modulators (agonists and antagonists) of expression and activity. The antibodies and antagonists may also be used as therapeutic agents to down regulate expression and activity. The antibodies may also be used as diagnostic agents for detecting the presence of the polypeptides in samples (e.g. by enzyme linked immunosorbent assay (ELISA). Examples of diseases which may be treated include rheumatoid arthritis and diabetes.
Sq	Sequence 2092 BP; 574 A; 390 C; 475 G; 653 T; 0 other;
Query Match	89.7%; Score 1824; DB 22; Length 2092;
Best Local Similarity	99.8%; Pred. No. 0;
Matches 2024; Conservative	0; Mismatches 4; Indels 0; Gaps 0;
Yy	1 gagtcgaaaggcgtactctctaagctggcggtgcgtccctgcccgcggccgcttc 60 
Dd	65 gagtcgaagagcgctcaacctccaacgcgtgcgcgtgccgtgccccgcgcgcgttc 124 
Qy	61 cgctcgcgccgcagctgctcgccgcgcgcgcgcgcacatgctgttgctgctgctctg 120 
Dd	125 cgtctgcgccccaatgctcgtcgccgcgcgcgcgcacatagctgtttgtcgtgctctg 184 
Qy	121 ggctttgctgcgcgggtgttttggggaacagctttagatatttaaatcacagaagtcttt 180 
Dd	185 ggcgttgtgctgcgggtgttttggggaacagattagatatttaaatcacagaagtcttt 244 
Qy	181 gttttcgaagaatggaattgctctaacaccagagagcggatccacaagctgtgcattg 240 
Dd	245 gttttcogaagaatggaattgctctaacaccagagagcggatccacaagctgtgcattg 304 
Qy	241 tcaatggctctctgtgaagaagaaccttcttggccagagctcgcagtgggtaacctg 300 
Dd	305 tcaatggctctctgtgaagaagaaccttcttggccagagctcgcagtggaacctg 364 

QY 301 ttctatcctccggtcgtacacgtcgtatggtgagggagtgaaacactggctcta 360  
 |||||||  
 Db 365 ttctatcctccggtcgtacacgtcgtatggtgagggagtgaaacactggctcta 424  
 QY 361 ccccaaggcagtgatcttctgtacaccttgaggagatgagtcctcttggttgaagt 420  
 |||||||  
 Db 425 ccccaaggcagtgatcttctgtacaccttgaggagatgagtcctcttggttgaagt 484  
 QY 421 gtgcgaattccatctactctta ttctctgaggaactccctgtgtttgcaagtggct 480  
 |||||||  
 Db 485 gtgcgaattccatctactctta ttctctgaggaactccctgtgtttgcaagtggct 544  
 QY 481 cccagtgaggaaaggtgtatagttagggagagcaactaagtggttgaagaccttca 540  
 |||||||  
 Db 545 cccagtgaggaaaggtgtatagttagggagagcaactaagtggttgaagaccttca 604  
 QY 541 gtcaactctgcagcgtccgttaacgcctgtttcaagaanaactgttctcaagttactc 600  
 |||||||  
 Db 605 gtcaactctgcagcgtccgttaacgcctgtttcaagaanaactgttctcaagttactc 664  
 QY 601 cccctcaattctctgagtaggaacaatgaggttgacctgctcttcttctgaactgca 660  
 |||||||  
 Db 665 cccctcaattctctgagtaggaacaatgaggttgacctgctcttcttctgaactgca 724  
 QY 661 gtgctacatgata ttcaagcttgctgtctgcataaagcatcagccaagagcatctct 720  
 |||||||  
 Db 725 gtgctacatgata ttcaagcttgctgtctgcataaagcatcagccaagagcatctct 784  
 QY 721 cctgattatattcactgtagagcttgaggtttgagatggaaagccttctgggaa 780  
 |||||||  
 Db 785 cctgattatattcactgtagagcttgaggtttgagatggaaagccttctgggaa 844  
 QY 781 gactctgaacaattcagagatgctcttaagaatcctgttgacgcctgcagaaagtttga 840  
 |||||||  
 Db 845 gactctgaacaattcagagatgctcttaagaatcctgttgacgcctgcagaaagtttga 904  
 QY 841 gattgagtagagcgtcttattatgttgagagacagtgtagatagtagcaactgtagca 900  
 |||||||  
 Db 905 gattgagtagagcgtcttattatgttgagagacagtgtagatagtagcaactgtagca 964  
 QY 901 ttggaacactccctcatctagagaagaagactatccttgaaggcaaaacagcgaaagac 960  
 |||||||  
 Db 965 ttggaacactccctcatctagagaagaagactatccttgaaggcaaaacagcgaaagac 1024  
 QY 961 ccaggaagctccatataccttgcatataagataatttgaalatctcgtgtttcaac 1020  
 |||||||  
 Db 1025 ccaggaagctccatataccttgcatataagataatttgaalatctcgtgtttcaac 1084  
 QY 1021 atgctacttgatgaatagatcgcttgcttgcttgcttgcttgatatacacttaacat 1080  
 |||||||  
 Db 1085 atgctacttgatgaatagatcgcttgcttgcttgcttgcttgatatacacttaacat 1144  
 QY 1081 tgggaacatgagatcctgtagatagatagcatcttattagagabtaacacagaaagattca 1140  
 |||||||  
 Db 1145 tgggaacatgagatcctgtagatagatagcatcttattagagabtaacacagaaagattca 1204  
 QY 1141 atgagatgaatgtagctgctgccaagaattagaagaagggtttggaattgctgttct 1200  
 |||||||  
 Db 1205 atgagatgaatgtagctgctgccaagaattagaagaagggtttggaattgctgttct 1264  
 QY 1201 taaatataatctttagtgtgcttaagaagtagatagataacttatacttaaaaaaa 1260  
 |||||||  
 Db 1265 taaatataatctttagtgtgcttaagaagtagatagataacttatacttaaaaaaa 1324  
 QY 1261 atcaaatgtgcttctattttgtgtgtgctgtgagttttcttagagtgaaattagt 1320  
 |||||||  
 Db 1325 atcaaatgtgcttctattttgtgtgtgctgtgagttttcttagagtgaaattagt 1384  
 QY 1321 attgagctgaatcccaactgtgtgtagatagatccataaatgcttgaaatattatgata 1380  
 |||||||  
 Db 1385 attgagctgaatcccaactgtgtgtagatagatccataaatgcttgaaatattatgata 1444

QY 1381 cattataacatgattccttcttctttaaagatlttggaatatgacgtgaagaagat 1440  
 Db 1445 cattataacatgattccttcttctttaaagatlttggaatatgacgtgaagaagat 1504  
 QY 1441 gtaaaacatttgaatagctcgtgttaltggaanaagtgacgtgaattattatgaacaac 1500  
 Db 1505 gtaaaacatttgaatagctcgtgttaltggaanaagtgacgtgaattattatgaacaac 1564  
 QY 1501 ttacgaatgcttaacttcttcaacagcataggtgaataatcatatttggcgatgtata 1560  
 Db 1565 ttacgaatgcttaacttcttcaacagcataggtgaataatcatatttggcgatgtata 1624  
 QY 1561 ctatgaacaattgttaaa tgccttaatttgatgtlaaataactcgtgaacaagaagaag 1620  
 Db 1625 ctatgaacaattgttaaa tgccttaatttgatgtlaaataactcgtgaacaagaagaag 1684  
 QY 1621 ttttaacttaagtagtacccttaaaatagatgtgcttataatgcgttagtttga 1680  
 Db 1685 ttttaacttaagtagtacccttaaaatagatgtgcttataatgcgttagtttga 1744  
 QY 1681 ctgatatcgtatagacagaggaagctgttltttaaaccctctcgcagtttgacct 1740  
 Db 1745 ctgatatcgtatagacagaggaagctgttltttaaaccctctcgcagtttgacct 1804  
 QY 1741 acatgggctataatgatactaaataactacatltgataaggaagaactgacctgtg 1800  
 Db 1805 acatgggctataatgatactaaataactacatltgataaggaagaactgacctgtg 1864  
 QY 1801 gatatataatgcttcttcatataacacacaaaatccctgaggaactlttgagacatg 1860  
 Db 1865 gatatataatgcttcttcatataacacacaaaatccctgaggaactlttgagacatg 1924  
 QY 1861 aataaaaaacatttatttctgaatcttccctgtgagtagtactatgttggag 1920  
 Db 1925 aataaaaaacatttatttctgaatcttccctgtgagtagtactatgttggag 1984  
 QY 1921 tacaactcatctatagataatataatgtaggaagtggtggaattctacttattatgtga 1980  
 Db 1985 tacaactcatctatagataatataatgtaggaagtggtggaattctacttattatgtga 2044  
 QY 1981 gtgagccaatgtctatcaagagtagcaaatagaagtaatgatgatctc 2028  
 Db 2045 gtgagccaatgtctatcaagagtagcaaatagaagtaatgatgatctc 2092

RESULT 5  
 AAZ40848  
 ID AAZ40848 standard; DNM; 2010 BP.  
 XX  
 AC AAZ40848;  
 XX  
 DT 18-JAN-2000 (first entry)  
 XX  
 DE Secreted protein EST coding sequence 33-77-4-E2-FUL.  
 XX  
 KW Secreted protein; fingerprint identification technique;  
 KW chromosome mapping; human; hereditary disease; diagnosis; cancer;  
 KW hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;  
 KW autoimmune disease; rheumatic disease; emphysema disorder; myopathy;  
 KW renal injury; amino aciduria; hypoglycaemia; male rat infertility;  
 KW hypertension; ss.  
 XX  
 OS Homo sapiens.  
 OS  
 PN M09940189-A2.  
 XX  
 PD 12-AUG-1999.  
 XX  
 PF 09-FEB-1999; 99MO-IB00282.  
 XX  
 PR 09-FEB-1998; 98US-0074121.  
 PR 13-APR-1998; 98US-0081563.  
 PR 10-AUG-1998; 98US-0096116.

PR 04-SEP-1996; 9805-009273.  
 PA (GENE) GENSET.  
 PI Bouqueleret L, Duclert A, Dumas Milne Edwards J:  
 XX WPI. 1999-600966/51.  
 DR P-PSDB; AAV59720.  
 XX

Extended cDNAs useful for expressing secreted proteins and to obtain specific antibodies -

Claim 1: Page 227-229; 244pp; English.

This sequence encodes a human secreted protein of the invention. The extended cDNAs (or genomic DNAs obtainable from them) may be used to prepare PCR primers and probes. These are useful for forensic matching or positive identification by DNA sequencing. They may also be used in alternative fingerprint identification techniques. Antibodies against the proteins encoded by the extended cDNAs are useful in identification of tissue types or cell species, as well as identifying tissue specific soluble proteins. The sequences can be used for chromosome mapping and identification of genes associated with hereditary diseases or drug response. Signal sequences from the cDNAs can be used in construction of secretion vectors. Other sequences derived from the extended cDNAs can be used to clone upstream genomic DNA sequences including promoters. This is in turn useful for identifying proteins that interact with promoter sequences. Some of the proteins may be useful in diagnosing and treating several disorders including, but not limited to: cancer, hyperlipidaemia, cardiovascular and neurodegenerative disorders, autoimmune diseases, and rheumatic diseases, embryonic disorders, hypertension, renal injury, amino acidurias, hypoglycaemia, male rat infertility and myopathies.

Sequence 2010 BP; 565 A; 360 C; 431 G; 642 T; 12 other;

Query Match 54.3%; Score 1103; DB 20; Length 2010;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1393; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 639 gctcttcttctgaacgagcgtgacatgataatgaagctgtgtctgcgtcaataa 698  
 DB 607 gctcttcttctgaacgagcgtgacatgataatgaagctgtgtctgcgtcaataa 666  
 QY 699 gcatctagccaagatcatctcctgattatatactgagcgtgaggtttgagca 758  
 DB 667 gcatctagccaagatcatctcctgattatatactgagcgtgaggtttgagca 726  
 QY 759 aattgggaagcgttattgggaagacctgaacaattcagagatgcttctaagatcctgt 818  
 DB 727 aattgggaagcgttattgggaagacctgaacaattcagagatgcttctaagatcctgt 786  
 QY 819 tgaagccttgaacaaatttgacagatgacatgacatgcttattatggtgggaagcgtgtgt 878  
 DB 787 tgaagccttgaacaaatttgacagatgacatgacatgcttattatggtgggaagcgtgtgt 846  
 QY 879 agaatgtactgtcgaagtcatttgacacctccctcattagaagaagaatactcct 938  
 DB 847 agaatgtactgtcgaagtcatttgacacctccctcattagaagaagaatactcct 906  
 QY 939 tgaagccttgaacaaatttgacagatgacatgacatgcttattatggtgggaagcgtgtgt 998  
 DB 907 tgaagccttgaacaaatttgacagatgacatgacatgcttattatggtgggaagcgtgtgt 966  
 QY 999 tgaatattcgtgtgtttcaacatggtactttgataatgacgcttgccttgcctgt 1058  
 DB 967 tgaatattcgtgtgtttcaacatggtactttgataatgacgcttgccttgcctgt 1026  
 QY 1059 gattatcactcttcaaatatttgaacaatg-attcctggagatagatagacattatata 1117  
 DB 1027 gattatcactcttcaaatatttgaacaatg-attcctggagatagatagacattatata 1086  
 QY 1118 ggtatgacaaacagaagatc-gaatgattgaaatgtaactgttgcctgacagaattagaag 1176

DB 1087 ggaatgacaaacagaagatc-gaatgattgaaatgtaactgttgcctgacagaattakaaag 1146  
 QY 1177 ggggttggaaattgctgtgttttgaataatacttctgtgtgtctttaaagtagatag 1236  
 DB 1147 ggggttggaaattgctgtgttttgaataatacttctgtgtgtctttaaagtagatag 1206  
 QY 1237 tatacttaacttataaaaaaaatcaaatgttcttattttttgtgtgtgtgtgtgtgtgtgt 1296  
 DB 1207 tatacttaacttataaaaaaaatcaaatgttcttattttttgtgtgtgtgtgtgtgtgtgt 1266  
 QY 1297 tgttttctgagtgaaatataagatgacgtgaatccactggtgataagattccataa 1356  
 DB 1267 tgttttctgagtgaaatataagatgacgtgaatccactggtgataagattccataa 1326  
 QY 1357 tatgttgaatattatgataagaccattataacattgatttattcctgtttatagat 1416  
 DB 1327 tatgttgaatattatgataagaccattataacattgatttattcctgtttatagat 1386  
 QY 1417 ttggaatattgacatgaaagaaatgtaaaacattagaatagctgtgtatgaaataa 1476  
 DB 1387 ttggaatattgacatgaaagaaatgtaaaacattagaatagctgtgtatgaaataa 1446  
 QY 1477 gtagcaggaattattagaacaaactgagatgagtttaacttcttatacagatagtgta 1536  
 DB 1447 gtagcaggaattattagaacaaactgagatgagtttaacttcttatacagatagtgta 1506  
 QY 1537 aaatcatatttgggtatgtatatactatgaaacatttgaatgttctaatttgaatga 1596  
 DB 1507 aaatcatatttgggtatgtatatactatgaaacatttgaatgttctaatttgaatga 1566  
 QY 1597 ataactctgaaacaaagaaaggttttaactagatagccttaataatgagatgtgtc 1656  
 DB 1567 ataactctgaaacaaagaaaggttttaactagatagccttaataatgagatgtgtc 1626  
 QY 1657 ttataatactgctgagtttggaaatgacgtgacatgaaacagagagagcgtgttttaac 1716  
 DB 1627 ttataatactgctgagtttggaaatgacgtgacatgaaacagagagagcgtgttttaac 1686  
 QY 1717 cctcttctgacagttgttgacacatgagcgtgacatgagcgtgacacacacacacacac 1776  
 DB 1687 cctcttctgacagttgttgacacatgagcgtgacatgagcgtgacacacacacacacac 1746  
 QY 1777 atctaaagaagaacacagccttggagatataagatgcttcaattacacacacacacacac 1836  
 DB 1747 atctaaagaagaacacagccttggagatataagatgcttcaattacacacacacacacac 1806  
 QY 1837 cccctgagagacatttggagcagatgaataataacacacacacacacacacacacacac 1896  
 DB 1807 cccctgagagacatttggagcagatgaataataacacacacacacacacacacacacac 1866  
 QY 1897 tgttgaatgactatggttgggtgtacacacacacacacacacacacacacacacacacac 1956  
 DB 1867 tgttgaatgactatggttgggtgtacacacacacacacacacacacacacacacacacac 1926  
 QY 1957 gtgaatcactacttattatggttggagcgaatgacacacacacacacacacacacacacac 2016  
 DB 1927 gtgaatcactacttattatggttggagcgaatgacacacacacacacacacacacacac 1986  
 QY 2017 aatgattgattccaaa 2032  
 DB 1987 aatgattgattccaaa 2002

RESULT 6  
 AAV49561  
 ID AAV49561 standard; cDNA to mRNA; 1050 BP.  
 XX  
 AC AAV49561;  
 XX  
 DT 21-OCT-1998 (first entry)  
 XX  
 DE Human epidermal carcinoma cell line KB clone HP01293 cDNA #2.

XX Transmembrane domain; human; nutrition; cytokine; cell proliferation;  
 KW differentiation; immune system; stimulator; suppressor; regulator;  
 KW haematopoiesis; activin; inhibitor; chemokine; chemokine; receptor;  
 KW haemostatic; thrombolytic; ligand; anti-inflammatory; tumour; ds.  
 OS Homo sapiens.  
 PN M09821328-A2.  
 XX  
 PD 22-MAY-1998.  
 XX  
 PF 07-NOV-1997; 97MO-JP04056.  
 XX  
 PR 13-NOV-1996; 96JP-0301429.  
 XX  
 PA (PROT-) PROTEGENE INC.  
 PA (SAGA) SAGAMI CHEM RES CENTRE.  
 XX  
 PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;  
 XX  
 DR WPI: 1998-297932/26.  
 DR P-PSDB; AAM64339.  
 XX  
 PT Human protein having transmembrane domain - useful for, e.g.  
 PT research and nutrition  
 XX  
 PS Claim 3; Page 121-122; 205pp; English.  
 XX  
 CC AA049550-V49599 are cDNA sequences which encode human proteins containing  
 CC a transmembrane domain. These proteins can be used for, e.g. research  
 CC and nutrition, and may have cytokine and cell  
 CC proliferation/differentiation, immune stimulating/suppressing,  
 CC haematopoiesis regulating, tissue growth, activin/inhibin,  
 CC chemotactic/chemokinetic, haemostatic and thrombolytic, receptor/ligand,  
 CC anti-inflammatory or tumour inhibition activity.  
 CC  
 XX  
 SQ Sequence 1050 BP; 261 A; 218 C; 253 G; 318 T; 0 other;

Query Match 51.6%; Score 1050; DB 19; Length 1050;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1050; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 97 atggctgtgttctgctcgttgcgttggtgaggtgtttgggaacgaattagt 156  
 |||||||  
 DB 1 atggctgtgttctgctcgttgcgttggtgaggtgtttgggaacgaattagt 60  
 OY 157 atataaatcaccaggtctgtgtttccgaatggaattgctatacagaagag 216  
 |||||||  
 DB 61 atataaatcaccaggtctgtgtttccgaatggaattgctatacagaagag 120  
 OY 217 cggatcccaagctgctcattgtcattggtctctcgtgaagaagaccttcttg 276  
 |||||||  
 DB 121 cggatcccaagctgctcattgtcattggtctctcgtgaagaagaccttcttg 180  
 OY 277 ccaagactcgaatgggtaacctgttcaatcctcctcgggtacacgtcgtgagtg 336  
 |||||||  
 DB 181 ccaagactcgaatgggtaacctgttcaatcctcctcgggtacacgtcgtgagtg 240  
 OY 337 aaggagtgaaacaactgctctacccccagagcgtatattctgtaaccttgaagat 396  
 |||||||  
 DB 241 aaggagtgaaacaactgctctacccccagagcgtatattctgtaaccttgaagat 300  
 OY 397 gcaattcctttagctctgaacagtggtgaattcaattcaattccttatttcttgagaa 456  
 |||||||  
 DB 301 gcaattcctttagctctgaacagtggtgaattcaattcaattccttatttcttgagaa 360  
 OY 457 actcctgtgttgaacagtggtcctcagtgagaaagatgtatagtaaggaaagca 516  
 |||||||  
 DB 361 actcctgtgttgaacagtggtcctcagtgagaaagatgtatagtaaggaaagca 420  
 OY 517 aactcagtggttgaagaaccttcaactccttgccagcctcgtlaactgcgttctca 576

DB 421 aactcagtggttgaagaaccttcaactccttgccagcctcgtlaactgcgttctca 480  
 |||||||  
 OY 577 gaaactcgttctcaactcctcctcctcctcctcctcctcctcctcctcctcctcctc 636  
 |||||||  
 DB 481 gaaactcgttctcagtcacccccctcctcctcctcctcctcctcctcctcctcctc 540  
 OY 637 ctgctcttcttctggaactcgaatgctcagatgatttcaagcttctcctcctcctc 696  
 |||||||  
 DB 541 ctgctcttcttctggaactcgaatgctcagatgatttcaagcttctcctcctcctc 600  
 OY 697 aagcatctagccaagatcattctcctgattatattcaactggaagctggaagcttgag 756  
 |||||||  
 DB 601 aagcatctagccaagatcattctcctgattatattcaactggaagctggaagcttgag 660  
 OY 757 gaattgggaagcgttatggggaagactctggaacattcagagatgcttcaagatcct 816  
 |||||||  
 DB 661 gaattgggaagcgttatggggaagactctggaacattcagagatgcttcaagatcct 720  
 OY 817 gtagcgtctgcgaagaatttgacagatgacatgacatgacatgacatgacatgacatg 876  
 |||||||  
 DB 721 gtagcgtctgcgaagaatttgacagatgacatgacatgacatgacatgacatgacatg 780  
 OY 877 gtagaattgacatgacatgacatgacatgacatgacatgacatgacatgacatgacatg 936  
 |||||||  
 DB 781 gtagaattgacatgacatgacatgacatgacatgacatgacatgacatgacatgacatg 840  
 OY 937 cttagagcaaaacgaagcgaagaacccaagacccaagacccaagacccaagacccaag 996  
 |||||||  
 DB 841 cttagagcaaaacgaagcgaagaacccaagacccaagacccaagacccaagacccaag 900  
 OY 997 tttaattcctggtgttttaacatgtaattggaatgataatgataatgataatgataatg 1056  
 |||||||  
 DB 901 tttaattcctggtgttttaacatgtaattggaatgataatgataatgataatgataatg 960  
 OY 1057 gtgatattcactcttaacatatttggaatgataatgataatgataatgataatgataat 1116  
 |||||||  
 DB 961 gtgatattcactcttaacatatttggaatgataatgataatgataatgataatgataat 1020  
 |||||||  
 OY 1117 aggatgacaacccaagatcgaatgat 1146  
 |||||||  
 DB 1021 aggatgacaacccaagatcgaatgat 1050  
 |||||||

RESULT 7  
 AA087120  
 ID AA087120 standard; cDNA; 577 BP.  
 XX  
 AC AA087120;  
 XX  
 DT 27-APR-1999 (first entry)  
 XX  
 DE EST clone B087.  
 XX  
 KW Expressed sequence tag; secreted protein; haematopoiesis regulator;  
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;  
 KW chemotaxis; chemokines; haemostasis; gene therapy; thrombolytic;  
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN M09845435-A2.  
 XX  
 PD 15-OCT-1998.  
 XX  
 PF 10-APR-1998; 98MO-US06954.  
 XX  
 PR 10-APR-1997; 97US-0835913.  
 XX  
 PA (GENM) GENETICS INST INC.  
 PA (GENM) GENETICS INST INC.  
 XX  
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;  
 PI Racie LA, Spaulding V, Treacy M;

XX WPI: 1999-070076/06.

XX New polynucleotides encoding human secreted proteins - derived from  
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
PT ovary, pituitary, retina and colon cDNA libraries

XX Claim 1: Page 459-460; 633pp; English.

XX This sequence represents an expressed sequence tag (EST), and is a  
CC polynucleotide of the invention. The polynucleotides of the invention are  
CC all secreted EST sequences isolated from a variety of human tissue  
CC sources. The EST sequences and proteins encoded by them are predicted to  
CC have useful biological activities which would make them suitable for  
CC treating, preventing or ameliorating medical conditions in humans and  
CC animals, although no supporting data is given. Suggested activities  
CC include nutritional activity, immune stimulating or suppressing activity,  
CC haematopoiesis regulating activity, tissue growth activity,  
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition  
CC activity. The EST sequences are also stated to be useful for gene  
CC therapy.

XX Sequence 577 BP; 195 A; 84 C; 102 G; 196 T; 0 other;

Query Match 24.3%; Score 495; DB 20; Length 577;

Best Local Similarity 99.8%; Pred. No. 8, 6e-229;

Matches 545; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1296 gttttctagatgaattatgattgacgtgaatccacgtgtagatagatccataat 1357  
DB 22 gttttctagatgaattatgattgacgtgaatccacgtgtagatagatccataat 81

QY 1358 atgctgaattatgatatagcattatgaattacattatctctgttaataaatt 1417

DB 82 atgctgaattatgatatagcattatgaattacattatctctgttaataaatt 141

QY 1418 tgggaataatgacatgaagaatgttaaacattagaatgctggttagaagaaga 1477

DB 142 tgggaataatgacatgaagaatgttaaacattagaatgctggttagaagaaga 201

QY 1478 tgcactgaattatgacagaactgaatgcttaacttttaacacgacatgagaa 1537

DB 202 tgcactgaattatgacagaactgaatgcttaacttttaacacgacatgagaa 261

QY 1538 aatataattgggcattatgatactatgaacaattgttaagtcttaattgagttaa 1597

DB 262 aatataattgggcattatgatactatgaacaattgttaagtcttaattgagttaa 321

QY 1598 taactctgaagaagaagaagtgtttaacttagagtagccctaaataatgagtgtc 1657

DB 322 taactctgaagaagaagaagtgtttaacttagagtagccctaaataatgagtgtc 381

QY 1658 tataataatgcttaattttggaactgatactgagtaacagagacagcgtttttaac 1717

DB 382 tataataatgcttaattttggaactgatactgagtaacagagacagcgtttttaac 441

QY 1718 ctctctgcaagtgttgacactacatggcctaataatgatactaaataactactatga 1777

DB 442 ctctctgcaagtgttgacactacatggcctaataatgatactaaataactactatga 501

QY 1778 tctaaagaagaactgctgtagatataatgattcattatcacacaataatc 1837

DB 502 tctaaagaagaactgctgtagatataatgattcattatcacacaataatc 561

QY 1838 cctgag 1843  
DB 562 cctgag 567

RESULT 8

AAF93970  
ID AAF93970 standard; DNA; 560 BP.

XX AAF93970;

XX 23-MAY-2001 (first entry)

DE Primer specific for DNA encoding secretory/membrane protein SEQ ID 404.

XX Human; secretory protein; membrane protein; vaccine; gene therapy;  
KW rheumatoid arthritis; diabetes; PCR primer; ss.

XX Synthetic.

XX EP1067182-A2.

XX 10-JAN-2001.

XX 07-JUL-2000; 2000EP-0114090.

XX 08-JUL-1999; 99JP-0194179.

XX 11-JAN-2000; 2000JP-0118775.

XX 02-MAY-2000; 2000JP-0183766.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;

XX WPI: 2001-093989/11.

XX Nucleic acids encoding secretory proteins/membrane proteins, useful in

XX gene therapy or as candidate target molecules in drug development -

XX Claim 4; SEQ ID 404; 609pp + CD ROM; English.

XX This invention relates to nucleic acid sequences AAF93744 - AAF93916

XX which encode human secretory or membrane proteins represented by

XX AAF88317 - AAF88419. Included in the invention are primers

XX AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the

XX cDNA sequences of the invention. The invention also includes methods for

XX the production of antibodies directed against the proteins, and cDNA

XX sequences, which can be used in vaccines. The polynucleotide sequences

XX can be used in gene therapy. The polynucleotide sequences and the

XX proteins they encode may be used in the prevention, treatment and

XX diagnosis of diseases associated with inappropriate secretory

XX protein/membrane protein expression. The nucleic acids and complementary

XX sequences may also be used as DNA probes in diagnostic assays

XX (e.g. polymerase chain reactions (PCR)) to detect and quantify the

XX presence of similar nucleic acid sequences in samples. They may also be

XX used to study the expression and function of secretory proteins/membrane

XX as antigens in the production of antibodies against them and in assays to

XX identify modulators (agonists and antagonists) of expression and

XX activity. The antibodies and antagonists may also be used as therapeutic

XX agents to down regulate expression and activity. The antibodies may also

XX be used as diagnostic agents for detecting the presence of the

XX polypeptides in samples (e.g. by enzyme linked immunosorbant assay

XX (ELISA). Examples of diseases which may be treated include rheumatoid

XX arthritis and diabetes.

XX Sequence 560 BP; 104 A; 145 C; 163 G; 142 T; 6 other;

Query Match 19.3%; Score 392; DB 22; Length 560;

Best Local Similarity 100.0%; Pred. No. 3, 9e-119;

Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gactccgagcgcgtcacctctcaacgctggtgctgcccgtgtccgcgcgcgcgttc 60

DB 66 gactccgagcgcgtcacctctcaacgctggtgctgcccgtgtccgcgcgcgcgttc 125

QY 61 cgtgcgccgcgcgtctgacgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 120

DB 126 cgtgccccagagtgctgcgcccgcgcgcacacatgctgtgttctgctccg 185  
 QY 121 gctgtgtgctgctgtgttctgtggaacagctttagatataataacacagctgtt 180  
 DB 186 gctgtgtgctgctgtgttctgtggaacagctttagatataataacacagctgtt 245  
 QY 181 gtttccgaatggaatgtgctacacagagagagagagagagagagagagagagag 240  
 DB 246 gtttccgaatggaatgtgctacacagagagagagagagagagagagagagagag 305  
 QY 241 tccatgagcttctgtgaaagaagacatttctgtgccaagagctgcagtgagtaacctg 300  
 DB 306 tccatgagcttctgtgaaagaagacatttctgtgccaagagctgcagtgagtaacctg 365  
 QY 301 tttatgtctctgctgctacacagctttagatgtgagagagagagagagagagagag 360  
 DB 366 tttatgtctctgctgctacacagctttagatgtgagagagagagagagagagagag 425  
 QY 361 cccccagcagtgctacattctgtacaccttga 392  
 DB 426 cccccagcagtgctacattctgtacaccttga 457

RESULT 9  
 AAF94119/c  
 ID AAF94119 standard; DNA; 588 BP.

XX AAF94119;

DT 23-MAY-2001 (first entry)

DE Primer specific for DNA encoding secretory/membrane protein SEQ ID 553.

KW Human; secretory protein; membrane protein; vaccine; gene therapy;  
 KW Rheumatoid arthritis; diabetes; PCR primer; ss.

OS Synthetic.

PN EPI067182-A2.

PD 10-JAN-2001.

XX 07-JUL-2000; 2000EP-0114090.

XX 08-JUL-1999; 99JP-0194179.

PR 11-JAN-2000; 2000JP-0118775.

PR 02-MAY-2000; 2000JP-0183766.

XX (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;

XX WPI: 2001-093989/11.

PT Nucleic acids encoding secretory proteins/membrane proteins, useful in  
 PT gene therapy or as candidate target molecules in drug development -

XX Claim 5; SEQ ID 553; 609BP + CD ROM; English.

XX This invention relates to nucleic acid sequences AAF93744 - AAF93916  
 CC which encode human secretory or membrane proteins represented by  
 CC AAB88317 - AAB88419. Included in the invention are primers  
 CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the  
 CC cDNA sequences of the invention. The invention also includes methods for  
 CC the production of antibodies directed against the proteins, and cDNA  
 CC sequences, which can be used in vaccines. The polynucleotide sequences  
 CC can be used in gene therapy. The polynucleotide sequences and the  
 CC proteins they encode may be used in the prevention, treatment and  
 CC diagnosis of diseases associated with inappropriate secretory  
 CC protein/membrane protein expression. The nucleic acids and complementary  
 CC sequences may also be used as DNA probes in diagnostic assays  
 CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the  
 CC presence of similar nucleic acid sequences in samples. They may also be

CC used to study the expression and function of secretory proteins/membrane  
 CC polypeptides and their role in metabolism. The polypeptides may be used  
 CC as antigens in the production of antibodies against them and in assays to  
 CC identify modulators (agonists and antagonists) of expression and  
 CC activity. The antibodies and antagonists may also be used as therapeutic  
 CC agents to down regulate expression and activity. The antibodies may also  
 CC be used as diagnostic agents for detecting the presence of the  
 CC polypeptides in samples (e.g. by enzyme linked immunosorbent assay  
 CC (ELISA)). Examples of diseases which may be treated include rheumatoid  
 CC arthritis and diabetes.  
 XX  
 SO Sequence 588 BP; 202 A; 114 C; 74 G; 193 T; 5 other;

Query Match 17.3%; Score 351; DB 22; Length 588;  
 Best Local Similarity 99.6%; Pred. No. 2.3e-159;  
 Matches 451; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1540 tcatattgggtatgtatctatctatgaacaattgtaaatgcttaattgtatgaata 1599  
 DB 489 tcatattgggtatgtatctatctatgaacaattgtaaatgcttaattgtatgaata 430  
 QY 1600 actctgaacaagaagaaggttttactactagtagccttaataatagatgtgtccta 1659  
 DB 429 actctgaacaagaagaaggttttactactagtagccttaataatagatgtgtccta 370  
 QY 1660 tataatcgtttagtttgggaactgtatctgagtaacagaagagagagagagagag 1719  
 DB 369 tataatcgtttagtttgggaactgtatctgagtaacagaagagagagagagagag 310  
 QY 1720 ctctcgaattgtgtgac 1779  
 DB 309 ctctcgaattgtgtgac 250  
 QY 1780 taagaagaacacagctctgtgagatataatagatgttctatataacacacacacac 1839  
 DB 249 taagaagaacacagctctgtgagatataatagatgttctatataacacacacacac 190  
 QY 1840 taagaagaacacagctctgtgagatataatagatgttctatataacacacacacac 1899  
 DB 189 taagaagaacacagctctgtgagatataatagatgttctatataacacacacacac 130  
 QY 1900 gtaagttaacatgt 1959  
 DB 129 gtaagttaacatgt 70  
 QY 1960 aatctacttttctatgttgagtgagacaaatgt 1992  
 DB 69 aatctacttttctatgttgagtgagacaaatgt 37

RESULT 10  
 AA217313  
 ID AA217313 standard; cDNA; 793 BP.

XX AA217313;

DT 12-OCT-1999 (first entry)

DE Human gene expression product cDNA sequence SEQ ID NO:4786.

XX Human; gene; gene expression product; diagnosis; therapy; probe;  
 KW detection; mapping; tissue typing; profiling; forensic; cancer;  
 KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

XX Homo sapiens.

XX WO938972-A2.

XX 05-AUG-1999.

XX 28-JAN-1999; 99WO-US01619.





families selected from 4 transmembrane segments integral membrane proteins, 7 transmembrane receptors, Arpases associated with various cellular activities (AAA), eukaryotic aspartyl proteases, GATA family of transcription factors, G-protein alpha subunit, photolipsters or diacylglycerol binding proteins, protein kinase, protein phosphatase 2C, protein tyrosine phosphatase, trypsin, wnt family of developmental signaling proteins and WW/rsp5/MMP domain containing proteins. The encoded polypeptides also have a functional domain selected from Ank repeat, basic region plus leucine zipper transcription factors, bromodomain, EF-hand, SH3 domain, WD domain/G-beta repeats, zinc finger (C2H2 type), zinc finger (CCHC class), and zinc-binding metalloprotease domain. The polynucleotides encode polypeptides with similarity to known protein families and are predicted to have similar properties. The novel polynucleotides can be used to develop products for use as therapeutic agents and in forensics, genetic analysis, mapping and diagnostic applications. In particular, the product can be used for the detection and management of cancers. They can be used for treating e.g. cervical cancers, melanomas, colorectal adenocarcinomas, Wilms' tumour, sarcomas, retinoblastoma, myosarcomas, lung carcinomas, leukemias, such as chronic myelogenous leukemia, promyelocytic leukemia, monocytic leukemia, and myeloid leukemia, and lymphomas such as histiocytic lymphoma, anhydric hereditary ectodermal dysplasia, congenital alveolar dysplasia, and epithelial dysplasia of the cervix, fibrous dysplasia of bone, and mammary dysplasia, hyperplasias, e.g. endometrial, adrenal, breast, prostate or thyroid hyperplasias or pseudoepitheliomatous hyperplasia of the skin.

Sequence 793 BP; 260 A; 108 C; 143 G; 267 T; 15 other;

Query Match 15.1%; Score 306; DB 20; Length 793;  
Best Local Similarity 99.3%; Pred. No. 1.1e-137;

Matches 576; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

1311 gaattatagatgacgtgaatccactggtgatgattccataatgcttgaatt 1370  
|||||  
58 gaattatagatgacgtgaatccactggtgatgattccataatgcttgaatt 117  
1371 atgatatgacatttaatacattgattcttcttgaattgaattgaattgac 1430  
|||||  
118 atgatatgacatttaatacattgattcttcttgaattgaattgaattgac 177  
1431 tgaagaagaatgaaacatttagaatgctgctgtaatgaaagaatgacgaatt 1490  
|||||  
178 tgaagaagaatgaaacatttagaatgctgctgtaatgaaagaatgacgaatt 237  
1491 ttagaacaattgcaatgcttaacttcttacaagcctagtggaatcattttgg 1550  
|||||  
238 ttanaacaacttgaagacttacttcttacaagcctagtg-aaacacatttgg 296  
1551 ctattgatacttgaacaattgttaaatgtcttaattgtaataacttgaaca 1610  
|||||  
297 ctattgatacttgaacaattgttaaatgtcttaattgtaataacttgaaca 356  
1611 aaagaagaagtttacttaacttagagagcctaaatgagtgcttataatgctt 1670  
|||||  
357 agagaagaagtttacttaacttagagagcctaaatgagtgcttataatgctt 416  
1671 agttttggaactgatactgatacagagagcagctgtttttaaaccctcttcga 1730  
|||||  
417 agttttggaactgatactgatacagagagcagctgtttttaaaccctcttcga 476  
1731 ttgttgacctacatgggctcaataatgatacaaaaataactacatgataagaagaac 1790  
|||||  
477 ttgttgacctacatgggctcaataatgatacaaaaataactacatgataagaagaac 536  
1791 taagcttggagatataatagatgcttcaatacaacacaaaatcccgaggagcatt 1850  
|||||  
537 taagcttggagatataatagatgcttcaatacaacacaaaatcccgaggagcatt 596  
1851 ttgagcatgataataacattttattcagtaattt 1890  
|||||  
597 ttgagcatgataataacattttattcagtaattt 636

# RESULT 12

AA215750  
ID AA215750 standard; cDNA; 813 BP.

XX  
XX AA215750:

DT 12-OCT-1999 (first entry)

DE Human gene expression product cDNA sequence SEQ ID NO:3219.

KW Human; gene; gene expression product; diagnosis; therapy; probe;

KW detection; mapping; tissue typing; profiling; forensic; cancer;

XX genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

OS Homo sapiens.

PN W09388972-A2.

PD 05-AUG-1999.

PE 28-JAN-1999; 99MO-US01619.

PR 03-APR-1998; 98US-0080666.

PR 28-JAN-1998; 98US-0072910.

PR 24-FEB-1998; 98US-0075954.

PR 31-MAR-1998; 98US-0080114.

PR 03-APR-1998; 98US-0080515.

XX  
XX PA (CHIR) CHIRON CORP.

PA (HYSE-) HYSFO INC.

PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;

PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;

PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;

PI Lamson G, Ieshkowitz D, Pot D, Randazzo F, Reinhard C;

PI Stache-Crain B, Sudduth-Klinger J, Williams LT;

XX  
XX WPI: 1999-494092/41.

XX  
XX Novel human genes and their expression products which are

PT differentially expressed in different cell types

XX  
XX Claim 1: Page 1545; 2479pp; English.

The present invention describes a library of human polynucleotides comprising the sequences given in AA21532 to AA21779. Also described is a method of detecting differentially expressed genes correlated with the cancerous state of a mammalian cell, comprising detecting at least one differentially expressed gene product in a test sample from a cell suspected of being cancerous, where the gene product is encoded by one of the 5248 polynucleotide sequences given in AA21532 to AA21779. The polynucleotides can be used as a source of primers and probes, which can be used for a variety of purpose, e.g. detection of expression levels, mapping, tissue typing or profiling, forensics, genetic analysis and detection of polymorphisms. Polypeptides encoded by the polynucleotides can be used for raising antibodies for experimental, diagnostic and therapeutic purposes. The polynucleotides may also be used to construct arrays for diagnostics (which may be used to determine function of an encoded protein); and to detect differences in expression levels between two cells (e.g. to identify abnormal or diseased tissue in a human, to identify a genetic predisposition or susceptibility to a disease such as cancer). The polynucleotides of the invention are especially used in the diagnosis, prognosis and management of colorectal cancer, breast cancer, and lung cancer. The polynucleotides can also be used to screen for peptide analogues and antagonists.

XX  
XX Sequence 813 BP; 251 A; 123 C; 143 G; 267 T; 29 other;

Query Match 13.1%; Score 266; DB 20; Length 813;

Best Local Similarity 100.0%; Pred. No. 2.2e-118;

Matches 266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1447 cattagaatgagcgtgtatgagaaagtgcaactgattatagacaactacga 1506  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 229 cattagaatgagcgtgtatgagaaagtgcaactgattatagacaactacga 288  
 QY 1507 atgcttaactctttacacagcatagtggaatcatcatttggtctattgatactatga 1566  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 289 atgcttaactctttacacagcatagtggaatcatcatttggtctattgatactatga 348  
 QY 1567 acaatttgaatgcttgaattgattgaataaacttgaacaaagaaagtttita 1626  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 349 acaatttgaatgcttgaattgattgaataaacttgaacaaagaaagtttita 408  
 QY 1627 acttagagtagccctaaatgatgagtgtctatataatcgcttagtttgcgaactgat 1686  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 409 acttagagtagccctaaatgatgagtgtctatataatcgcttagtttgcgaactgat 468  
 QY 1687 ctgaatcaacagagacagctgtttt 1712  
 ||||||||||||||||||||||||||||  
 Db 469 ctgaatcaacagagacagctgtttt 494

## RESULT 13

AAZ1759  
 ID AAZ1759 standard; cDNA; 813 BP.

AC AAZ1759;

DT 12-OCT-1999 (first entry)

DE Human gene expression product cDNA sequence SEQ ID NO:3228.

KW Human; gene; gene expression product; diagnosis; therapy; probe; detection; mapping; tissue typing; profiling; forensic; cancer; genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

OS Homo sapiens.

PN W09938972-A2.

XX 05-AUG-1999.

PF 28-JAN-1999; 99WO-US01619.

PR 03-APR-1998; 98US-0080666.

PR 28-JAN-1998; 98US-0072910.

PR 24-FEB-1998; 98US-0075954.

PR 31-MAR-1998; 98US-0080114.

PR 03-APR-1998; 98US-0080515.

PA (CHIR ) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

PI Cirvenjakov R, Dickson M, Drmanac R, Drmanac S, Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA, Jones WL, Kassam A, Kennedy GC, Kita D, Labat I, Lamson G, Lesnikowitz D, Pot D, Randazzo F, Reinhard C, Stache-Grain B, Sudduth-Klinger J, Williams LT;

PI MPI; 1999-494092/41.

PT Novel human genes and their expression products which are differentially expressed in different cell types

PS Claim 1; Page 1549; 2479pp; English.

XX The present invention describes a library of human polynucleotides comprising the sequences given in AAZ12532 to AAZ17779. Also described is a method of detecting differentially expressed genes correlated with the cancerous state of a mammalian cell, comprising detecting at least one differentially expressed gene product in a test sample from a cell suspected of being cancerous, where the gene product is encoded by one

CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The CC polynucleotides can be used as a source of primers and probes, which can be used for a variety of purpose, e.g. detection of expression levels, CC mapping, tissue typing or profiling, forensics, genetic analysis and CC detection of polymorphisms. Polypeptides encoded by the polynucleotides CC can be used for raising antibodies for experimental, diagnostic and CC therapeutic purposes. The polynucleotides may also be used to construct CC arrays for diagnostics (which may be used to determine function of an CC encoded protein); and to detect differences in expression levels between CC two cells (e.g. to identify abnormal or diseased tissue in a human, to CC identify a genetic predisposition or susceptibility to a disease such as CC cancer). The polynucleotides of the invention are especially used in the CC diagnosis, prognosis and management of colorectal cancer, breast cancer, CC and lung cancer. The polynucleotides can also be used to screen for CC peptide analogues and antagonists.

SO Sequence 813 BP; 251 A; 123 C; 143 G; 267 T; 29 other;

Query Match 13.1%; Score 266; DB 20; Length 813;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-118;  
 Matches 266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1447 cattagaatgagcgtgtatgagaaagtgcaactgattatagacaactacga 1506  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 229 cattagaatgagcgtgtatgagaaagtgcaactgattatagacaactacga 288  
 QY 1507 atgcttaactctttacacagcatagtggaatcatcatttggtctattgatactatga 1566  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 289 atgcttaactctttacacagcatagtggaatcatcatttggtctattgatactatga 348  
 QY 1567 acaatttgaatgcttgaattgattgaataaacttgaacaaagaaagtttita 1626  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 349 acaatttgaatgcttgaattgattgaataaacttgaacaaagaaagtttita 408  
 QY 1627 acttagagtagccctaaatgatgagtgtctatataatcgcttagtttgcgaactgat 1686  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 409 acttagagtagccctaaatgatgagtgtctatataatcgcttagtttgcgaactgat 468  
 QY 1687 ctgaatcaacagagacagctgtttt 1712  
 ||||||||||||||||||||||||||||  
 Db 469 ctgaatcaacagagacagctgtttt 494

## RESULT 14

AAZ17715  
 ID AAZ17715 standard; cDNA; 416 BP.

AC AAZ17715;

DT 12-OCT-1999 (first entry)

DE Human gene expression product cDNA sequence SEQ ID NO:5188.

KW Human; gene; gene expression product; diagnosis; therapy; probe; detection; mapping; tissue typing; profiling; forensic; cancer; genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

OS Homo sapiens.

PN W09938972-A2.

XX 05-AUG-1999.

PF 28-JAN-1999; 99WO-US01619.

PR 03-APR-1998; 98US-0080666.

PR 28-JAN-1998; 98US-0072910.

PR 24-FEB-1998; 98US-0075954.

PR 31-MAR-1998; 98US-0080114.

PA (CHIR ) CHIRON CORP.



QY 1969 ttctatgttgagtgagccaatgtctatcaagatgacaaataaagttaatgatgtcc 2028  
 |||  
 Db 194 ttctatgttgagtgagccaatgtctatcaagatgacaaataaagttaatgatgtcc 253  
 QY 2029 aaaa 2031  
 |||  
 Db 254 aaaa 256

## RESULT 16

AAK51918  
 ID AAK51918 standard; DNA; 343 BP.

AC AAK51918;

DT 22-JUN-1999 (first entry)

XX Human secreted protein 5' EST SEQ ID NO: 132.

XX Human; secreted protein; EST; expressed sequence tag; diagnosis;  
 KM forensic; gene therapy; chromosome mapping; signal peptide;  
 KM upstream regulatory sequence; cytokine activity; cell proliferation;  
 KM differentiation; haematopoiesis regulation; tissue growth regulation;  
 KM reproductive hormone regulation; chemokine; chemokine; haemostatic;  
 KM thrombolytic; anti-inflammatory; tumour inhibition; ds.

OS Homo sapiens.

PN WO9906552-A2.

PD 11-FEB-1999.

PF 31-JUL-1998; 98WO-1B01236.

PR 01-AUG-1997; 97US-0905223.

PA (GEST ) GENSET.

PI Ducleert A, Dumas Milne Edwards J, Lacroix B;

DR WPI: 1999-153782/13.  
 P-PSDB: AAY13118.

PT New isolated brain-derived nucleic acids - used to develop products  
 PT which may have cytokine, immune, regulatory, haematopoiesis  
 PT regulating, anti-inflammatory or tumour inhibition activity

PS Claim 1; Page 273-274; 577pp; English.

CC AAK51787 to AAK52019 represent 5' expressed sequence tags (ESTs) for  
 CC human secreted proteins, and encode the proteins given in AAY12987 to  
 CC AAY13119, respectively. The proteins given represent the signal peptide  
 CC and an N-terminal fragment of a secreted protein. The nucleic acid  
 CC sequences can be used for producing secreted human gene products. They  
 CC can also be used to develop products for diagnosis and therapy. The  
 CC proteins obtained may have cytokine activity, cell  
 CC proliferation/differentiation activity, haematopoiesis regulating  
 CC activity, tissue growth regulating activity, reproductive hormone  
 CC regulating activity, chemotactic/chemokinetic activity, haemostatic and  
 CC thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, tumour inhibition activity or other activities. The products  
 CC can be used in forensic, gene therapy and chromosome mapping procedures.  
 CC The sequences can also be used for obtaining corresponding promoter  
 CC sequences. The nucleic acids encoding of a polypeptide can be used for  
 CC directing extracellular secretion of a polypeptide or the insertion of a  
 CC polypeptide into a membrane, or importing a polypeptide into a cell.

CC Sequence 343 BP; 56 A; 88 C; 104 G; 88 T; 7 other;

Query Match 11.2%; Score 227; DB 20; Length 343;

Best Local Similarity 100.0%; Pred. No. 14e-99;  
 Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 caccatgctgtgttctgtcgtcctcgtgagcgttctgtgagcgttctgtgagcgttct 152  
 |||  
 Db 61 caccatgctgtgttctgtcgtcctcgtgagcgttctgtgagcgttctgtgagcgttct 120  
 QY 153 tagtatataaatacaccagagctgtgttcttcgaaatgaaatgagcctataccagg 212  
 |||  
 Db 121 tagtatataaatacaccagagctgtgttcttcgaaatgaaatgagcctataccagg 180  
 QY 213 agagcggatccaccagcgtgtgtcgtatgtccatgagcgttcttcgtgaagaagaccttc 272  
 |||  
 Db 181 agagcggatccaccagcgtgtgtcgtatgtccatgagcgttcttcgtgaagaagaccttc 240  
 QY 273 ttgagccagactcgcagtggtgaacctgttaccatgcctcctcgagccta 319  
 |||  
 Db 241 ttgagccagactcgcagtggtgaacctgttaccatgcctcctcgagccta 287

## RESULT 17

AAZ13071  
 ID AAZ13071 standard; cDNA; 297 BP.

AC AAZ13071;

DT 12-OCT-1999 (first entry)

XX Human gene expression product cDNA sequence SEQ ID NO:540.

DE Human; gene; gene expression product; diagnosis; therapy; probe;

KW detection; mapping; tissue typing; profiling; forensic; cancer;  
 KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

OS Homo sapiens.

PN WO9938972-A2.

PD 05-AUG-1999.

PF 28-JAN-1999; 99WO-US01619.

PR 03-APR-1998; 98US-0080666.

PR 28-JAN-1998; 98US-0072910.

PR 24-FEB-1998; 98US-0075954.

PR 31-MAR-1998; 98US-0080114.

PR 03-APR-1998; 98US-0080515.

PA (CHIR ) CHIRON CORP.

PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;  
 PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;  
 PI Lamson G, Lesnikowitz D, Pot D, Randazzo F, Reinhard C;  
 PI Stache-Crain B, Sudduth-Klinger J, Williams LV;

DR WPI: 1999-494092/41.

PT Novel human genes and their expression products which are  
 PT differentially expressed in different cell types

PS Claim 1; Page 784-785; 2479pp; English.

CC The present invention describes a library of human polynucleotides  
 CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is  
 CC a method of detecting differentially expressed genes correlated with the  
 CC cancerous state of a mammalian cell, comprising detecting at least one  
 CC differentially expressed gene product in a test sample from a cell  
 CC suspected of being cancerous, where the gene product is encoded by one  
 CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The  
 CC polynucleotides can be used as a source of primers and probes, which can  
 CC be used for a variety of purpose, e.g. detection of expression levels,  
 CC mapping, tissue typing or profiling, forensics, genetic analysis and  
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides





PT New nucleic acid fragment encoding gene products - can be used  
 for genetic analysis and mapping  
 XX  
 PS Claim 1; Page 92; 616pp; English.  
 XX  
 CC Human nucleic acid fragments, isolated from brain adrenal tissue,  
 CC the placenta or bone marrow comprise any of: (A) a sequence  
 CC selected from (AA076401-Q77613), (B) an allelic variation of a  
 CC sequence as described in (A), or (C) a sequence complementary  
 CC to (A) or (B).  
 CC Preferred sequences exhibit no more than 90% homology to a human  
 CC sequence known per se.  
 XX  
 SQ Sequence 204 BP; 58 A; 18 C; 41 G; 81 T; 6 other;

Query Match 5.7%; Score 116; DB 15; Length 204;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-46;  
 Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DY 1145 attgaatgacctgctgcgaattagaagaagggttggaattgctgtttgttaa 1204  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 16 attgaatgacctgctgcgaattagaagaagggttggaattgctgtttgttaa 75  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DY 1205 atatacttttagtgccttaagaatagatagatattacattataaaaaa 1260  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 76 atatacttttagtgccttaagaatagatagatattacattataaaaaa 131  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 22  
 AA076548  
 ID AA076548 standard; DNA; 148 BP.  
 XX  
 AC AA076548;  
 XX  
 XX 23-SEP-1994 (first entry)  
 XX  
 DE Human genome fragment.  
 XX  
 KW Brain; placenta; bone marrow; genetic analysis; gene mapping;  
 KW detection; homology; human; adrenal tissue; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9401548-A.  
 XX  
 PD 20-JAN-1994.  
 XX  
 PF 13-JUL-1993; 93MO-GB01467.  
 XX  
 PR 13-JUL-1992; 92GB-0014857.  
 XX  
 PA (MEDICAL RES COUNCIL.  
 XX  
 PI Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;  
 PI Stibson DR, Starkey M;  
 DR WPI: 1994-035056/04.  
 XX  
 PT New nucleic acid fragment encoding gene products - can be used  
 PT for genetic analysis and mapping  
 XX  
 PS Claim 1; Page 91; 616pp; English.  
 XX  
 CC Human nucleic acid fragments, isolated from brain adrenal tissue,  
 CC the placenta or bone marrow comprise any of: (A) a sequence  
 CC selected from (AA076401-Q77613), (B) an allelic variation of a  
 CC sequence as described in (A), or (C) a sequence complementary  
 CC to (A) or (B).  
 CC Preferred sequences exhibit no more than 90% homology to a human  
 CC sequence known per se.  
 XX  
 SQ Sequence 148 BP; 47 A; 15 C; 24 G; 57 T; 5 other;

Query Match 4.9%; Score 99; DB 15; Length 148;  
 Best Local Similarity 100.0%; Pred. No. 7.5e-38;  
 Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DY 1162 ccagaattgaaagggttggaattgctgtttgttaaatatattatcttagt 1221  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 31 ccagaattgaaagggttggaattgctgtttgttaaatatattatcttagt 90  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DY 1222 cttaaatagatagatattacattataaaaaa 1260  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 91 cttaaatagatagatattacattataaaaaa 129  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 23  
 AA215908  
 ID AA215908 standard; CDNA; 828 BP.  
 XX  
 AC AA215908;  
 XX  
 XX 12-OCT-1999 (first entry)  
 XX  
 DE Human gene expression product cDNA sequence SEQ ID NO:3377.  
 XX  
 KW Human; gene; gene expression product; diagnosis; therapy; probe;  
 KW detection; mapping; tissue typing; profiling; forensic; cancer;  
 KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9938972-A2.  
 XX  
 PD 05-AUG-1999.  
 XX  
 PF 28-JAN-1999; 99MO-US01619.  
 XX  
 PR 03-APR-1998; 98US-0080666.  
 PR 28-JAN-1998; 98US-0072910.  
 PR 24-FEB-1998; 98US-0075954.  
 PR 31-MAR-1998; 98US-0080114.  
 PR 03-APR-1998; 98US-0080515.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (HYSE-) HYSE INC.  
 XX  
 PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;  
 PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;  
 PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;  
 PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;  
 PI Stache-Crain B, Sudduth-Klinger J, Williams LT;  
 DR WPI: 1999-494092/41.  
 XX  
 PT Novel human genes and their expression products which are  
 PT differentially expressed in different cell types  
 XX  
 PS Claim 1; Page 1617; 2479pp; English.  
 XX  
 CC The present invention describes a library of human polynucleotides  
 CC comprising the sequences given in AA21532 to AA21779. Also described is  
 CC a method of detecting differentially expressed genes correlated with the  
 CC cancerous state of a mammalian cell, comprising detecting at least one  
 CC differentially expressed gene product in a test sample from a cell  
 CC suspected of being cancerous, where the gene product is encoded by one  
 CC of the 5248 polynucleotide sequences given in AA21532 to AA21779. The  
 CC polynucleotides can be used as a source of primers and probes, which can  
 CC be used for a variety of purpose, e.g. detection of expression levels,  
 CC mapping, tissue typing or profiling, forensics, genetic analysis and  
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides  
 CC can be used for raising antibodies for experimental, diagnostic and  
 CC therapeutic purposes. The polynucleotides may also be used to construct  
 CC arrays for diagnostics (which may be used to determine function of an

CC encoded protein); and to detect differences in expression levels between  
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to  
CC identify a genetic predisposition or susceptibility to a disease such as  
CC cancer). The polynucleotides of the invention are especially used in the  
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,  
CC and lung cancer. The polynucleotides can also be used to screen for  
CC peptide analogues and antagonists.

XX Sequence 828 BP; 204 A; 132 C; 134 G; 230 T; 128 other;

Query Match 1.0%; Score 20; DB 20; Length 828;  
Best Local Similarity 100.0%; Pred. No. 9;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1652 tttgcttataatcgctta 1671  
Db 398 tttgcttataatcgctta 417

RESULT 24

AAC40298  
ID AAC40298 standard; DNM; 1174 BP.

XX AAC40298;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 27803.

XX Hybridisation assay; genetic mapping; gene expression control;  
XX protein identification; signal transduction pathway;

KM metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

PN EPI033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132464.

PR 05-MAY-1999; 99US-0132485.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 14-JUN-1999; 99US-0138847.

PR 16-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 17-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 21-JUN-1999; 99US-0139753.

PR 22-JUN-1999; 99US-0139817.

PR 23-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 24-JUN-1999; 99US-0140354.

PR 28-JUN-1999; 99US-0140695.

PR 29-JUN-1999; 99US-0140823.

PR 30-JUN-1999; 99US-0140991.

PR 01-JUL-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 02-JUL-1999; 99US-0142154.

PR 06-JUL-1999; 99US-0142055.

PR 08-JUL-1999; 99US-0142390.

PR 09-JUL-1999; 99US-0142803.

PR 12-JUL-1999; 99US-0142920.

PR 13-JUL-1999; 99US-0142977.

PR 14-JUL-1999; 99US-0143542.

PR 15-JUL-1999; 99US-0143624.

PR 16-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145086.

PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.



```

PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159333.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160777.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.

```

```

PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match 1.0%; Score 20; DB 21; Length 1174;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 2012 aagtaaatgatgatccaa 2031
Db 296 aagtaaatgatgatccaa 315

```

## RESULT 25

```

AAZ56976/c
ID AAZ56976 standard; cDNA; 1718 BP.

```

```

AAZ56976;

```

```

AC 12-MAY-2000 (first entry)

```

```

DE Arabidopsis aldehyde dehydrogenase (ALDH)-4 encoding cDNA.

```

```

XX Arabidopsis: plant plastid; acetyl CoA synthetase; E3 subunit; ACS;
KW plant plastidic pyruvate dehydrogenase; pPDH; ATP citrate lyase; ACL;
KW pyruvate decarboxylase; PDC; aldehyde dehydrogenase; ALDH; acetyl CoA;
KW fatty acid; flavonoid; enzyme; phytochemical; pyruvate decarboxylase;
KW acetyl CoA hydrolase; mitochondrial pyruvate dehydrogenase; ss.

```

```

XX Arabidopsis sp.

```

```

XX WO20000619-A2.

```

```

PD 06-JAN-2000.

```

```

XX 25-JUN-1999; 99WO-US14382.

```

```

PF 26-JUN-1998; 98US-0090717.

```

```

PR (IOWA ) UNIV IOWA STATE RES FOUND INC.

```

```

PA Nikolau BJ, Wurtele ES, Oliver DJ, Behal R, Schnable PS, Ke J;
PI Johnson JL, Allred CC, Faltand B, Lutziger I, Wen T;

```

```

XX WPI: 2000-160678/14.

```

```

DR P-PSDB; AAY67415.

```

```

PT Novel acetyl CoA synthetase (ACS), plastidic pyruvate dehydrogenase
PT (pPDH), ATP citrate lyase (ACL), pyruvate decarboxylase (PDC) and
PT aldehyde dehydrogenase (ALDH) polypeptides used to alter acetyl CoA
PT levels in plants

```

```

PS Claim 31; Fig 13; 79pp; English.

```

```

XX The invention provides nucleic acids encoding Arabidopsis plastidic
CC acetyl CoA synthetase (ACS), various subunits (specifically the E3
CC subunit) of plant plastidic pyruvate dehydrogenase (pPDH), the A and B
CC subunits of a plant ATP citrate lyase (ACL), Arabidopsis pyruvate
CC decarboxylase (PDC), Arabidopsis aldehyde dehydrogenase (ALDH),
CC specifically ALDH-2 and ALDH-4. The polypeptides can be expressed by
CC standard recombinant methodology. The ACS, pPDH, ACL, PDC and ALDH
CC polypeptides, methods and nucleic acid molecules of the invention are
CC or organs. A decrease in acetyl CoA is expected to affect the
CC biosynthesis of very long chain fatty acids and flavonoids. The enzymes
CC may also be used for in vitro synthesis of acetyl CoA, which in turn can
CC be used to produce acetyl CoA phytochemicals (plastidic ACS, pPDH, ACL,
CC pyruvate decarboxylase, acetyl CoA hydrolase, mitochondrial pyruvate
CC dehydrogenase and aldehyde dehydrogenase are involved in the biosynthesis
CC of acetyl CoA.

```

```

SQ Sequence 1718 BP; 528 A; 344 C; 380 G; 466 T; 0 other;

```

Query Match 1.0%; Score 20; DB 21; Length 1718;  
 Best Local Similarity 100.0%; Pred. No. 8.9;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1431 tgaagaatgtaaacatt 1450  
 |||  
 Db 1660 TGAAGAATGTAACATT 1641

RESULT 26  
 AAQ41241/C  
 ID AAQ41241 standard; mRNA, 172 BP.

AC AAQ41241;

XX 26-AUG-1993 (first entry)

DE 3'-5' m1cf RNA strand (hybridises to ompF mRNA 5'-3').

KW Stem loop structures; gene expression; drug resistance; cancer;  
 KM m1cf: mRNA interfering complementary RNA; osmoregulation;  
 KW bidirectional; outer membrane protein; E. coli; ds.

XX Escherichia coli.

Key Location/Qualifiers

FT stem\_loop 1..21

FT /tag= a

FT /note= "stem loop c of m1cf RNA"

FT stem\_loop 23..49

FT /tag= b

FT /note= "stem loop b of m1cf RNA"

FT misc\_structure 62..66

FT /tag= c

FT /note= "loop"

FT misc\_structure 75..79

FT /tag= d

FT /note= "loop"

FT misc\_structure 93..98

FT /tag= e

FT /note= "loop"

FT stem\_loop 144..166

FT /tag= f

FT /note= "stem loop a of m1cf RNA"

US5208149-A.

PD 04-MAY-1993.

XX 20-OCT-1983; 83US-0543528.

XX 20-OCT-1983; 83US-0543528.

XX 01-MAR-1984; 84US-0585282.

XX 03-AUG-1988; 88US-0228852.

XX 23-JAN-1989; 89US-0300741.

XX 15-NOV-1989; 89US-0436598.

XX 29-MAY-1990; 90US-0530159.

XX 10-APR-1992; 92US-0870186.

XX (UYNV) UNTV NEW YORK STATE RES FOUND.

XX Inouye M;

XX WPI; 1993-159156/19.

XX Nucleic acid constructs contain stable stem and loop structures -

XX and mRNA interfering complementary RNA, for regulating gene

XX expression, e.g. oncogene(s) of viral genes

XX Disclosure: Fig 3; 26pp; English.

XX Gene expression of the E. coli outer membrane proteins ompC and ompF

CC is osmoregulated. The ompC locus is transcribed bidirectionally under  
 CC conditions of high osmolarity. The upstream stretch of mRNA of ca.  
 CC 170 bp (CX28) inhibits prodn. of ompF protein. This mRNA (m1cf RNA)  
 CC has a sequence complementary to the 5' end of ompF mRNA and hybridises  
 CC to prevent gene expression. Very stable stem loop structures a, b  
 CC and c form within the repeat regions of m1cf RNA. These structures  
 CC are characteristic of rho-factor independent transcription termination  
 CC sites in prokaryotes. The hybrid between ompF RNA and CX28 RNA is  
 CC sandwiched between the stable stem loop structures. This mechanism  
 CC may be adapted for other nucleotide sequences, to regulate, inhibit  
 CC and/or control gene expression. These systems may be used to alter  
 CC drug resistance genes, viral genes, etc. In the treatment of genetic  
 CC defects and diseases, cancers, etc.  
 CC See also AAQ41240-6.

CC Sequence 172 BP; 51 A; 39 C; 19 G; 63 U; 0 other;

Query Match 0.9%; Score 19; DB 14; Length 172;  
 Best Local Similarity 100.0%; Pred. No. 28;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2007 aaataaagttaagtat 2025  
 |||  
 Db 102 AAATAAGTTAATGATGAT 84

RESULT 27

AAQ20669/C  
 ID AAQ20669 standard; RNA; 174 BP.

AC AAQ20669;

XX 16-APR-1992 (first entry)

DE m1cf RNA.

XX mRNA interfering complementary RNA system; artificial; ss.

XX Synthetic.

XX Key Location/Qualifiers

FT misc\_binding 58..139

FT /tag= a

FT /note= "binds to ompF gene region"

FT misc\_feature 24..50

FT /tag= b

FT /note= "5' end stem-and-loop structure"

FT terminator 146..168

FT /tag= c

FT /note= "3' end rho-independent transcription

FT termination signal"

EP467349-A.

PD 22-JAN-1992.

XX 19-OCT-1984; 84EP-0111983.

XX 17-JUL-1991; 91EP-0111983.

XX 20-OCT-1983; 83US-0543528.

XX 01-MAR-1984; 84US-0585282.

XX (UYNV) UNTV OF N YORK.

XX Inouye M, Mizuno T, Chou MY;

XX WPI; 1992-026189/04.

XX Non-human organism contg. nucleic acid construct - regulates gene

XX function and produces ribonucleotide sequence

XX Disclosure: Fig 3; 26pp; English.

XX The sequence is that of artificial mRNA interfering complementary  
 CC (micf) which is complementary to a part of the ompF gene transcript  
 CC and forms an extremely stable hybrid with it. 44 bases of the 5' end  
 CC untranslated region of ompF mRNA, including the Shine-Dalgarno  
 CC sequence and 28 bases from the coding region are involved in the  
 CC hybrid formation. This hybrid structure is occurs between two stable  
 CC stem-and-loop structures of the micF RNA, one for the 3' end rho-  
 CC independent transcription termination signal and the other at the 3'  
 CC end. It can be used to allow regulation of expression of the ompF  
 CC gene in E. coli, the system using mic RNA can be used to regulate  
 CC expression of any specific gene in E. coli. The inducible lethality  
 CC which can be created using this system may also be useful for studying  
 CC essential genes. It can also be used to block the expression of harmful  
 CC genes, e.g. oncogenes and viral genes, and to influence the expression  
 CC of other genes.

SQ Sequence 174 BP; 51 A; 40 C; 19 G; 64 U; 0 other:

Query Match 0.9%; Score 19; DB 13; Length 174;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2007 aaataagtaatgatgat 2025  
 |||||

DB 103 AAATAAGTTAATGATGAT 85

RESULT 28  
 AAQ53309/C  
 ID AAQ53309 standard; RNA; 174 BP.

AC AAQ53309;

XX 20-JUN-1994 (first entry)

DE MicF-RNA complementary to region of ompF mRNA.

XX Outer membrane protein; ompC; ompF; promoter; gene expression;

KW osmoregulation; gene regulation; complementary sequence;

KW hybridisation; antisense; Escherichia coli; ss.

XX Escherichia coli.

XX key Location/Qualifiers

FT stem\_loop 1..22

FT stem\_loop 24..50

FT stem\_loop 146..168

FT stem\_loop 146..168

FT stem\_loop 146..168

FT stem\_loop 146..168

FT stem\_loop 146..168

FT stem\_loop 146..168

FT stem\_loop 146..168

FT stem\_loop 146..168

FT stem\_loop 146..168

FT stem\_loop 146..168

FT stem\_loop 146..168

FT stem\_loop 146..168

XX WPI: 1993-413402/51.

XX Regulating gene function in a cell - using poly:nucleotide  
 CC construct which produces poly:nucleotide complementary to RNA  
 CC transcript of the gene

XX Disclosure; Figure 3; 25pp; English.

XX A stretch of DNA approximately 300 bp long located upstream of the  
 CC ompC promoter is able to block ompF expression. This region (CX28  
 CC fragment) contains a transcription unit which is directed towards  
 CC the region upstream of the ompC gene so that mic-F, the RNA  
 CC transcript from this region has a sequence complementary to the ompF  
 CC mRNA. The hybridisation between the two RNA's blocks the  
 CC translation of ompF mRNA to ompF protein. Using this as a model  
 CC system it is possible to devise a method of regulating the function  
 CC of a gene in a cell comprising introducing into a cell a non-native  
 CC polynucleotide construct which produces a non-naturally occurring  
 CC polynucleotide complementary to an RNA transcript produced by the  
 CC gene and can then regulate the function of the gene.

SQ Sequence 174 BP; 51 A; 40 C; 19 G; 64 U; 0 other:

Query Match 0.9%; Score 19; DB 14; Length 174;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2007 aaataagtaatgatgat 2025  
 |||||

DB 103 AAATAAGTTAATGATGAT 85

RESULT 29  
 AAQ37973/C  
 ID AAQ37973 standard; DNA; 174 BP.

AC AAQ37973;

XX 30-JUN-1993 (first entry)

DE micF-RNA.

XX gene expression; regulate; control; inhibit; enhance; increase;

KW decrease; block expression; oncogenes; viral genes.

XX E. coli.

XX key Location/Qualifiers

FT stem\_loop 1..22

FT stem\_loop 62..66

FT stem\_loop 76..80

FT stem\_loop 76..80

FT stem\_loop 76..80

FT stem\_loop 76..80

FT stem\_loop 76..80

FT stem\_loop 76..80

FT stem\_loop 76..80

FT stem\_loop 76..80

FT stem\_loop 76..80

FT stem\_loop 76..80

US5190931-A.

PR 02-MAR-1993.

PR 15-NOV-1989;

PR 20-OCT-1983;

PR 03-AUG-1988;

83US-0543528.

84US-0585282.

88US-0228852.

89US-0436598.

89US-0543528.

84US-0585282.

88US-0228852.

89US-0436598.

89US-0543528.

83US-0543528.

84US-0585282.

88US-0228852.

89US-0436598.

89US-0543528.

83US-0543528.

84US-0585282.

88US-0228852.

89US-0436598.

89US-0543528.

83US-0543528.

84US-0585282.

88US-0228852.

89US-0436598.

89US-0543528.

83US-0543528.

84US-0585282.

88US-0228852.

89US-0436598.

89US-0543528.

83US-0543528.

84US-0585282.

88US-0228852.

89US-0436598.

89US-0543528.

PR 23-JAN-1989; 89US-0300741.  
PR 15-NOV-1989; 89US-0436598.  
XX  
XX (UYNY ) UNIV NEW YORK STATE RES FOUND  
PA  
XX

Query Match	0.98;	Score 19;	DB 14;	Length 174;
Best Local Similarity	100.0%;	Pred. No. 28;		
Matches 19;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	2007	aaataaagttaatgatgat	2025
Db	103	AAATAAGTTAATGATGAT	85

RESULT	30
AAAX21425	
ID	AAAX21425 standard; DNA; 420 BP
vv	

CC proteins especially non-native plasmid or plant proteins. This sequence  
CC corresponds to the regulatory region for the plasmid *Ip*adA(1).  
XX  
SQ Sequence 420 BP; 120 A; 55 C; 60 G; 185 T; 0 other;

Query Match	0.98;	Score 19;	DB 20;	Length 420;
Best Local Similarity	100.0%;	Pred. No. 28;		
Matches 19;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 1253 aaaaaaaaaatcaaatlttg 1271  
 |||||  
 Db 299 aaaaaaaaaatcaaatlttg 317

RESULT	31
AAN50239	
ID	AAN50239 standard; DNA; 500 BP.
XY	

PI Inouye M, Mizuno T, Chou MY;  
 XX  
 DR WPI: 1985-111906/19.  
 XX  
 PT Regulation of gene expression in organisms - by using  
 PT translational inhibition with mRNA-interfering complementary RNA  
 XX  
 PS Disclosure; Fig 2: 57pp; English.  
 XX  
 CC ompf and ompC are osmoregulation genes. The ompC locus was found to  
 CC be transcribed bidirectionally under conditions of high osmolality,  
 CC and the upstream transcript RNA of approximately 170 bases was found  
 CC to inhibit the production of OmpF protein. This RNA (micRNA) has a  
 CC long sequence which is complementary to the 5'-end region of the  
 CC ompf mRNA that includes the ribosome-binding site and the coding  
 CC region of the first nine AA residues of pro-OmpF protein. Thus, it  
 CC is proposed that micRNA inhibits the translation of ompf mRNA by  
 CC hybridizing with it.  
 XX  
 SO Sequence 500 BP; 174 A; 59 C; 108 G; 159 T; 0 other:

Query Match 0.9%; Score 19; DB 6; Length 500;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2007 aataaagtaatgatgat 2025  
 ||||||||||||||||  
 DB 135 aataaagtaatgatgat 153

RESULT 32  
 ID AAO53308 standard; DNA; 500 BP.  
 XX  
 AC AAO53308;  
 XX  
 DT 20-JUN-1994 (first entry)  
 XX  
 DE Promoter and upstream region of E. Coli ompC gene.  
 XX  
 KM Outer membrane protein; ompC; ompF; promoter; gene expression;  
 KM osmoregulation; gene regulation; complementary sequence;  
 KM hybridisation; antisense; Escherichia coli; ds.  
 OS Escherichia coli.  
 XX  
 FH Key Location/Qualifiers  
 FT -10\_signal Complement (245..250)  
 FT -35\_signal /\*tag= a  
 FT Complement (267..272)  
 FT /\*tag= b  
 FT -35\_signal 377..382  
 FT /\*tag= c  
 FT -10\_signal 400..405  
 FT /\*tag= d  
 XX  
 XX US5272065-A.  
 PN  
 PD 21-DEC-1993.  
 XX  
 PE 20-OCT-1983; 83US-0543528.  
 XX  
 PR 20-OCT-1983; 83US-0543528.  
 PR 01-MAR-1984; 84US-0585282.  
 PR 03-AUG-1988; 88US-0228852.  
 PR 23-JAN-1989; 89US-0300741.  
 PR 15-NOV-1989; 89US-0436598.  
 PR 21-JUN-1990; 90US-0542209.  
 XX  
 PA (UNYX ) UNIV NEW YORK STATE RES FOUND.  
 XX  
 PI Chou M, Inouye M, Mizuno T;

XX  
 DR WPI: 1993-413402/51.  
 XX  
 PT Regulating gene function in a cell - using poly:nucleotide  
 PT construct which produces poly:nucleotide complementary to RNA  
 PT transcript of the gene  
 XX  
 PS Disclosure; Figure 2: 25pp; English.  
 XX  
 CC A stretch of DNA approximately 300 bp long located upstream of the  
 CC ompC promoter is able to block ompF expression. This region (CX28  
 CC fragment) contains a transcription unit which is directed towards  
 CC the region upstream of the ompC gene so that the RNA transcript  
 CC from this region has a sequence complementary to the ompf mRNA. The  
 CC hybridisation between the two RNA's blocks the translation of ompf  
 CC mRNA to OmpF protein. Using this as a model system it is possible  
 CC to devise a method of regulating the function of a gene in a cell  
 CC comprising introducing into a cell a non-natively polynucleotide  
 CC construct which produces a non-naturally occurring polynucleotide  
 CC complementary to an RNA transcript produced by the gene and can then  
 CC regulate the function of the gene.  
 XX  
 SO Sequence 500 BP; 174 A; 59 C; 108 G; 159 T; 0 other:

Query Match 0.9%; Score 19; DB 14; Length 500;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2007 aataaagtaatgatgat 2025  
 ||||||||||||||||  
 DB 135 aataaagtaatgatgat 153

RESULT 33  
 ID AAO41240 standard; DNA; 500 BP.  
 XX  
 AC AAO41240;  
 XX  
 DT 26-AUG-1993 (first entry)  
 XX  
 DE Promoter and upstream region of OmpC.  
 XX  
 KM Stem loop structures; gene expression; drug resistance; cancer;  
 KM micF; mRNA interfering complementary RNA; osmoregulation;  
 KM bidirectional; outer membrane protein; E. coli; ds.  
 XX  
 OS Escherichia coli.  
 XX  
 FH Key Location/Qualifiers  
 FT repeat\_region /\*tag= a  
 FT 70..92  
 FT repeat\_unit 70..81  
 FT /\*tag= b  
 FT /note= "inverted repeat unit"  
 FT repeat\_region 187..214  
 FT /\*tag= c  
 FT repeat\_unit 187..200  
 FT /\*tag= d  
 FT /note= "inverted repeat unit"  
 FT repeat\_region 216..237  
 FT /\*tag= e  
 FT repeat\_unit 216..227  
 FT /note= "inverted repeat unit"  
 FT -10\_signal Complement (245..250)  
 FT /\*tag= g  
 FT /note= "micF promoter"  
 FT Complement (267..272)  
 FT -35\_signal /\*tag= h  
 FT /note= "micF promoter"  
 FT Complement (293..300)  
 FT misc\_feature



CC This sequence shows 500bp from the XbaI site to the initiation  
CC codon AUG of the ompC gene of E. coli. The DNA downstream of the  
CC residue 88 was determined previously. It was found that from  
CC residue 99 to 180 had 70% homology with the 5' end of the ompF mRNA  
CC which includes the Shine-Dalgarno sequence, the initiation codon,  
CC and the codons for the first nine amino acids of pro-ompF protein. A  
CC plausible model to explain this result is that the 300bp CX28 fragment  
CC contains a transcription unit which is directed towards the region  
CC upstream of the ompC gene so that the RNA transcript from this region  
CC has a sequence complementary to the ompF mRNA. The hybridisation between  
CC the two RNAs thus blocks the translation of ompF mRNA to OmpF protein.  
XX  
SQ Sequence 500 BP; 174 A; 59 C; 108 G; 159 T; 0 other;

Query Match 0.9%; Score 19; DB 14; Length 500;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2007 aataaagtaatgatgat 2025  
|||||  
DB 135 aataaagtaatgatgat 153

## RESULT 35

AAZ22886  
ID AAZ22886 standard; DNA; 625 BP.

AC AAZ22886;

XX 20-DEC-1999 (first entry)

DE E. coli mutant ompC gene.

XX Bacteria: attenuation; deletion; mutant; vaccine; immune response;

KW Gram negative; infection; diarrhoea; food poisoning; typhoid;

KW salmonellosis; gonorrhoea; gastroenteritis; whooping cough; ds.

OS Escherichia coli.

OS Synthetic.

PN W09949026-A1.

PD 30-SEP-1999.

PF 25-MAR-1999; 99WO-GB00935.

PR 25-MAR-1998; 98GB-0006449.

PA (PEPT-) PEPTIDE THERAPEUTICS LTD.

PI Chatfield SN;

DR WPI: 1999-580447/49.

PT New attenuated bacteria useful as vaccines for protecting against

XX Infections

PS Example 1; Page -: 69pp; English.

CC This sequence represents a E. coli mutant ompC gene. The coding sequence  
CC of the wild-type gene (AAZ22885) was removed via PCR using primers TT7-  
CC TT10 (AAZ22894-222896) to produce a non-reverting deletion mutation.  
CC This sequence was used in the production of a bacterium attenuated by a  
CC non-reverting mutation in each of the ompC gene, the aroC gene  
CC (AAZ22883), and the ompF gene (AAZ22887, AAZ22888). The mutant  
CC bacteria provide immunogenic activity with reduced virulence and thus can  
CC be used as a vaccine for raising an immune response against a variety of  
CC bacteria in a mammalian host. Such vaccines can provide protection  
CC against e.g., E. coli (a cause of diarrhoea in humans), Salmonella  
CC typhimurium (the cause of salmonellosis in several animal species), S.  
CC typhi (the cause of human typhoid), S. enteritidis (a cause of food  
CC poisoning in humans), S. choleraesuis (a cause of salmonellosis in pigs),

CC S. dublin (a cause of both a systemic and diarrhoeal disease in cattle,  
CC especially of new-born calves), Haemophilus influenzae (a cause of  
CC meningitis), Neisseria gonorrhoeae (a cause of gonorrhoea), Yersinia  
CC enterocolitica (the cause of a spectrum of disease in humans ranging  
CC from gastroenteritis to fatal septicemic disease), Bordetella pertussis  
CC (the cause of whooping cough), and Brucella abortus (a cause of abortion  
CC and infertility in cattle and a condition known as undulant fever in  
CC humans).  
CC Note: The present sequence is not shown in the specification but is  
CC derived from the E. coli aroC gene sequence given on pages 48-50.  
XX  
SQ Sequence 625 BP; 201 A; 89 C; 137 G; 198 T; 0 other;

Query Match 0.9%; Score 19; DB 20; Length 625;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2007 aataaagtaatgatgat 2025  
|||||  
DB 135 aataaagtaatgatgat 153

## RESULT 36

AAZ99064  
ID AAZ99064 standard; cDNA; 953 BP.

AC AAZ99064;

XX 24-SEP-1999 (first entry)

DE Human validated cancer cell derived cDNA #386.

XX Cancer: human; colon; breast; lung; transmembrane receptor; ATPase;

KW integral membrane protein; aspartyl protease; GATA family; wnt family;

KW transcription factor; G-protein alpha subunit; protein phosphatase;

KW phospholipase binding protein; diacylglycerol binding protein; trypsin;

KW protein kinase; tyrosine phosphatase; developmental signalling protein;

KW WW/rps25 domain; therapy; forensic; genetic mapping; diagnostic;

KW detection; treatment; cervical; melanoma; colorectal adenocarcinoma;

KW Wilms' tumour; retinoblastoma; sarcoma; myosarcoma; lung carcinoma;

KW leukemia; lymphoma; dysplasia; hyperplasia; endometrium; adrenal;

XX prostate; ss.

OS Homo sapiens.

OS W09933982-A2.

PN 08-JUL-1999.

PD 22-DEC-1998; 98WO-US27610.

PF 21-DEC-1998; 98US-0217471.

PR 23-DEC-1997; 97US-0068755.

PR 03-APR-1998; 98US-0080664.

PR 21-OCT-1998; 98US-0105234.

PR 27-OCT-1998; 98US-0105877.

XX (CHIR) CHIRON CORP.

PA (HYSE-) HYSEO INC.

PI Cirvenjakov R, Dickson M, Drmanac R, Drmanac S;

PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;

PI Jones LW, Kassam A, Kennedy GC, Kita D, Labat I;

PI Lanson G, Desikowitz D, Pot D, Randazzo F, Reinhard C;

PI Stache-Crain B, Sudduth-Klinger J, Williams LT;

DR WPI: 1999-430243/36.

XX New isolated human polynucleotides

PT Claim 1; Page 568; 591pp; English.

XX

CC This invention describes novel isolated human polynucleotides obtained  
 CC by screening for differential expression in colon cancer, breast cancer  
 CC and lung cancer cell lines. The polynucleotides of the invention are  
 CC represented in AX98275-X99118 and encode polypeptides of protein  
 CC families selected from 4 transmembrane segments integral membrane  
 CC proteins, 7 transmembrane receptors, ATRases associated with various  
 CC cellular activities (AMAs), eukaryotic aspartyl proteases, GATA family of  
 CC transcription factors, G-protein alpha subunit, phospholipases or  
 CC diacylglycerol binding proteins, protein kinase, protein phosphatase 2C,  
 CC protein tyrosine phosphatase, trypsin, wnt family of developmental  
 CC signalling proteins and WW/trsp/MMP domain containing proteins. The  
 CC encoded polypeptides also have a functional domain selected from Ank  
 CC repeat, basic region plus leucine zipper transcription factors,  
 CC bromodomain, EF-hand, SH3 domain, WD domain/G-beta repeats, zinc finger  
 CC (C2H2 type), zinc finger (C2HC class), and zinc-binding metalloprotease  
 CC domain. The polynucleotides encode polypeptides with similarity to known  
 CC protein families and are predicted to have similar properties. The novel  
 CC polynucleotides can be used to develop products for use as therapeutic  
 CC agents and in forensics, genetic analysis, mapping and diagnostic  
 CC applications. In particular, the product can be used for the detection  
 CC and management of cancers. They can be used for treating e.g. cervical  
 CC cancers, melanomas, colorectal adenocarcinomas, Wilms' tumour, sarcomas,  
 CC retinoblastoma, myosarcomas, lung carcinomas, leukemias, such as chronic  
 CC myelogenous leukemia, promyelocytic leukemia, monocytic leukemia, and  
 CC myeloid leukemia, and lymphomas such as histiocytic lymphoma, anaplastic  
 CC hereditary ectodermal dysplasia, congenital alveolar dysplasia,  
 CC epithelial dysplasia of the cervix, fibrous dysplasia of bone, and  
 CC mammary dysplasia, hyperplasias, e.g. endometrial, adrenal, breast,  
 CC prostate or thyroid hyperplasias or pseudoeplitheliomatous hyperplasia of  
 CC the skin.

CC Sequence 953 BP; 255 A; 164 C; 150 G; 182 T; 202 other;

Query Match 0.9%; Score 19; DB 20; Length 953;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1253 aaaaaaatcaatttg 1271  
 ||||||||||||||||  
 Db 639 aaaaaaatcaatttg 657

RESULT 37

AXX21432  
 ID AAX21432 standard; DNA: 1416 BP.

AC AAX21432;

DT 21-MAY-1999 (first entry)

DE Regulatory region Prrn(L)/rbcl(S)/Kan/TpsbA(L).

KW Construct; marker; antibiotic resistance; regulatory sequence; promoter;  
 KW stabilizing sequence; plasmid; plant; ss.

OS Synthetic.

PN US5877402-A.

PD 02-MAR-1999.

PF 31-JAN-1994; 94US-0189256.

PR 31-JAN-1994; 94US-0189256.

PR 01-MAY-1990; 90US-0518763.

PR 25-AUG-1993; 93US-0111398.

PA (RUTE ) UNIV RUTGERS STATE NEW JERSEY.

PI Allison LA, Carrer H, Kanevski I, Maliqa P, Maliqa ZS;  
 PI Staub JM, Zoubenko O;

XX

DR WPI; 1999-189744/16.  
 XX  
 PT New DNA construct containing an antibiotic resistance marker -  
 PT useful for stably transforming the plasmids of multicellular plants  
 XX  
 PS Disclosure; Fig 28A; 101pp; English.

CC The invention relates to a DNA construct, containing a non-lethal stable  
 CC marker (e.g. antibiotic resistance) under the control of a 5' regulatory  
 CC sequence and a 3' stabilizing sequence, for stably transforming the  
 CC plasmids of multicellular plants and allowing expression of heterologous  
 CC proteins especially non-native plasmid or plant proteins. This sequence  
 CC corresponds to the regulatory region Prrn(L)/rbcl(S)/Kan/TpsbA(L) from  
 CC the plasmid pTmH7.

CC Sequence 1416 BP; 312 A; 322 C; 379 G; 403 T; 0 other;

Query Match 0.9%; Score 19; DB 20; Length 1416;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1253 aaaaaaatcaatttg 1271  
 ||||||||||||||||  
 Db 1295 aaaaaaatcaatttg 1313

RESULT 38

AAT85196  
 ID AAT85196 standard; DNA: 1417 BP.

AC AAT85196;

DT 10-MAR-1998 (first entry)

DE Plasmid targeting region of plasmid pGS85A.

KW Plasmid; transformation; transplastomic plant; transgenic plant;

KW Brassica; cruciferous plant; vector; plasmid pGS85A;

KW kanamycin resistance; neomycin phosphotransferase; neo gene;

OS Chimeric - Arabidopsis thaliana.

OS Chimeric - Synthetic.

FT Key Location/Qualifiers  
 FT CDS 158..970  
 FT /\*tag= a  
 FT /note= "kanamycin resistance gene"

PN WO9732977-A1.

PD 12-SEP-1997.

PF 06-MAR-1997; 97WO-US03444.

PR 06-MAR-1996; 96US-0012916.

PA (RUTE ) UNIV RUTGERS STATE NEW JERSEY.

PI Maliqa P, Reddy SV, Sikdar SR;

PI WPI; 1997-470543/43.

PT Production of transplastomic plants - by transfection with plasmid  
 PT targeted DNA, specifically for Cruciferous plants

PS Disclosure; Fig 7; 45pp; English.

CC This sequence comprises the plasmid targeting region of plasmid  
 CC pGS85A. This plasmid carries a neomycin phosphotransferase (neo)  
 CC gene that confers kanamycin resistance gene, flanked on both sides  
 CC by Arabidopsis plasmid targeting sequences. It was produced by



ligating the neo gene into the unique HincII site of plasmid pGS7 (see Atr85194). A novel method of producing transplasmic plants involves delivering transforming DNA, preferably cloned in pGS3A, pGS83A or pGS7, to a plastid genome of plant cells in culture. Selecting cells with transformed plastids and regenerating these to plants. Transforming DNA comprises (i) (1) several targeting sequences (i.e. plastid DNA sequences from the genome to be transformed) to allow homologous recombination, (ii) 5' and 3' regulatory regions from plastid DNA linked to a selectable marker gene, (iii) regulatory regions controlling a foreign gene of interest, and (iv) at a cloning site for insertion of the foreign gene adjacent to the marker gene. The method is used to transform circulatorous plants, e.g. Brassica species. The combination of homologous recombination, selectable marker and use of plastid regulatory sequences improves generation of stably transformed plants. Expression of foreign genes in plastids, rather than in the nucleus, avoids risk of transmitting the foreign gene in the pollen, provides high levels of protein expression, permits simultaneous incorporation of several genes and avoids the positional effects and gene silencing associated with nuclear transformation.

Sequence 1417 BP; 312 A; 323 C; 379 G; 403 T; 0 other:

Query Match 0.9%; Score 19; DB 18; Length 1417;

Best Local Similarity 100.0%; Pred. No. 27; Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1253 aaaaaaatcaatttg 1271  
|||||  
DB 1296 aaaaaaatcaatttg 1314

RESULT 39

AAC38836/C  
ID AAC38836 standard; DNA; 1700 BP.

XX AAC38836;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 22408.

KM Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;

KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

XX EP103405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 16-APR-1999; 99US-0128714.

PR 19-APR-1999; 99US-0129845.

PR 21-APR-1999; 99US-0130077.

PR 23-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132683.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136382.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139753.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.

PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 16-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144632.

PR 21-JUL-1999; 99US-0144684.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 26-JUL-1999; 99US-0145224.  
 PR 27-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 28-JUL-1999; 99US-0145919.  
 PR 02-AUG-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 03-AUG-1999; 99US-0146389.  
 PR 04-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 05-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 06-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 09-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 10-AUG-1999; 99US-0147935.  
 PR 11-AUG-1999; 99US-0148171.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0150854.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 30-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.

PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161922.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 0.9%; Score 19; DB 21; Length 1700;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1646 tatgatgcttataataa 1664  
 ||||||||||||||||  
 Db 1628 TATGATGCTCTATATATAA 1610

## RESULT 40

AAZ22885  
 ID AAZ22885 standard; DNA; 1713 BP.

AC AAZ22885;  
 XX

DT 20-DEC-1999 (first entry)  
 XX

DE E. coli wild-type ompC gene.  
 XX

KW Bacteria; attenuation; deletion; mutant; vaccine; immune response;  
 KW Gram negative; infection; diarrhoea; food poisoning; typhoid;  
 KW salmonellosis; gonorrhoea; gastroenteritis; whooping cough; ds.

OS Escherichia coli.  
 XX

FT Key Location/Qualifiers  
 FT CDS 492..1562  
 FT /\*tag= a  
 FT /product= "E. coli wild-type ompC protein"

PN W09949026-A1.  
 XX

PD 30-SEP-1999.  
 XX

PP 25-MAR-1999; 99WO-GR00935.  
 XX

PR 25-MAR-1998; 98GB-0006449.  
 XX

PA (PEPT-) PEPTIDE THERAPEUTICS LTD.  
 XX

PI Chatfield SN;  
 XX

DR WPI: 1999-580447/49.  
 XX

DK P-PSDB; AAI42549.  
 XX

PT New attenuated bacteria useful as vaccines for protecting against  
 PT infections -  
 XX

PS Disclosure; Page 53-56; 69pp; English.  
 XX

CC This sequence represents the E. coli wild-type ompC gene. The coding  
 CC sequence was removed via PCR using primers TT7-TT10 (AAZ22893-22896)  
 CC to produce a non-reverting deletion mutation. The mutant ompC gene  
 CC (AAZ22886) was used in the production of a bacterium attenuated by a  
 CC non-reverting mutation in each of the ompC gene, the atoc gene

CC (AA222883, AA222884), and the ompF gene (AA222887, AA222888). The mutant  
 CC bacteria provide immunogenic activity with reduced virulence and thus can  
 CC be used as a vaccine for raising an immune response against a variety of  
 CC bacteria in a mammalian host. Such vaccines can provide protection  
 CC against e.g., E. coli (a cause of diarrhoea in humans), Salmonella  
 CC typhimurium (the cause of salmonellosis in several animal species), S.  
 CC typhi (the cause of human typhoid), S. enteritidis (a cause of food  
 CC poisoning in humans), S. choleraesuis (a cause of salmonellosis in pigs),  
 CC S. dublin (a cause of both a systemic and diarrhoeal disease in cattle,  
 CC especially of new-born calves), Haemophilus influenzae (a cause of  
 CC meningitis), Neisseria gonorrhoeae (a cause of gonorrhoea), Yersinia  
 CC enterocolitica (the cause of a spectrum of disease in humans ranging  
 CC from gastroenteritis to fatal septicemic disease), Bordetella pertussis  
 CC (the cause of whooping cough), and Brucella abortus (a cause of abortion  
 CC and infertility in cattle and a condition known as undulant fever in  
 CC humans).

SO Sequence 1713 BP; 481 A; 373 C; 404 G; 455 T; 0 other;

Query Match 0.9%; Score 19; DB 20; Length 1713;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2007 aaataaagtaatgatgat 2025  
 |||  
 DB 135 aaataaagtaatgatgat 153

RESULT 41  
 AAA30825  
 ID AAA30825 standard; DNA; 4196 BP.

XX AAA30825;  
 XX  
 DT 29-AUG-2000 (first entry)  
 XX

DE Expression cassette pGEN2.

XX Expression cassette; plasmid maintenance system; Neisseria meningitidis;  
 KW post-segregational killing function; ompC promoter; immune response;  
 KW vaccine; Salmonella typhi; hepatitis; Haemophilus influenzae type b;  
 KW acellular pertussis; varicella; rotavirus; Streptococcus pneumoniae;  
 KW cancer vaccine; autoimmune disorder; immunological disease; allergy;  
 KW myasthenia gravis; lupus erythematosus; rheumatoid arthritis; therapy;  
 KW multiple sclerosis; asthma; pGEN2; ss.

XX Synthetic.

OS  
 XX WO200032047-A1.

PN 08-JUN-2000.

PD 02-DEC-1999; 99WO-US28499.

PF 02-DEC-1998; 98US-0204117.

PR 12-OCT-1999; 99US-0158738.

XX (UYMA-) UNIV MARYLAND BALTIMORE.

PA Galen JE;

PI WPI; 2000-412091/35.

XX Expression cassette used as live vector vaccine comprises nucleotide  
 PT sequence encoding origin of replication and plasmid maintenance system  
 PT which includes a post-segregational killing and a partitioning function

XX Example; Fig 4; 127pp; English.

XX This sequence represents an example of an expression cassette of the  
 CC invention, designated pGEN2. The cassette is an independently

CC functioning expression cassette (I), comprises a nucleotide sequence  
 CC encoding an origin of replication (ORI) and a nucleotide sequence  
 CC encoding a plasmid maintenance system (PMS) which includes a  
 CC post-segregational killing function (PSK) and a partitioning function. An  
 CC expression plasmid containing a modified ompC promoter is used for  
 CC eliciting an immune response in a human or bovine subject. (I) is used  
 CC for transforming a bacterial cell which is used to elicit an immune  
 CC response. (I) can also be used to vaccinate a subject against Salmonella  
 CC typhi. (I) may comprise an antigen for hepatitis B, Haemophilus  
 CC influenzae type b, hepatitis A, acellular pertussis (act), varicella,  
 CC rotavirus, Streptococcus pneumoniae, or Neisseria meningitidis, and can  
 CC be used as vaccines against diseases caused by these agents. (I) can be  
 CC also used as a cancer vaccine. The antigens encoded by the plasmids are  
 CC designed to provoke an immune response to autoantigens, B cell receptors  
 CC and/or T cell receptors which are implicated in autoimmune or  
 CC immunological diseases, where an inappropriate immune response is raised  
 CC against body tissues, or environmental antigens, the vaccines may  
 CC immunise against the autoantigens, B cell receptors and/or T cell  
 CC receptors to modulate the responses and ameliorate diseases, such as  
 CC myasthenia gravis, lupus erythematosus, rheumatoid arthritis, multiple  
 CC sclerosis, allergies and asthma. The PMS incorporated into multicopy  
 CC expression plasmids encoding one or more proteins or peptides of  
 CC interest, enhances the level of expression of the protein or peptide of  
 CC interest. The PMS provides improved stability of recombinant plasmids,  
 CC overcoming prior art problems of plasmid instability.

SO Sequence 4196 BP; 1084 A; 1082 C; 1062 G; 968 T; 0 other;

Query Match 0.9%; Score 19; DB 21; Length 4196;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2007 aaataaagtaatgatgat 2025  
 |||  
 DB 118 aaataaagtaatgatgat 136

RESULT 42

AA261385/C  
 ID AA261385 standard; DNA; 4671 BP.

XX AA261385;

AC 19-JUN-2000 (first entry)

DT Nucleotide sequence of plasmid pMSK35.

XX aada: protein expression; vaccine; haemoglobin; enzyme;

KW green fluorescent protein; ss.

XX Synthetic.

OS Key Location/Qualifiers

FT misc\_feature 1..1880 /tag= a

FT misc\_feature 1917..3249 /note= "rice left targeting sequence"

FT misc\_feature 3285.. /tag= b

FT misc\_feature 1758..4671 /note= "green fluorescent protein region"

FT /tag= d /note= "rice right targeting sequence"

PN WO200007431-A1.

PD 17-FEB-2000.

PF 03-AUG-1999; 99WO-US17806.

PR 03-AUG-1998; 98US-0095163.  
 PR 03-AUG-1998; 98US-0095167.  
 PR 15-DEC-1998; 98US-0112257.  
 PR 29-APR-1999; 99US-0131611.  
 PR 11-JUN-1999; 99US-0138764.  
 XX

PA (RUTF ) UNIV ROTTERS STATE NEW JERSEY.

PI Maliga P, Kuroda H, Khan MS;

XX WPI; 2000-205525/18.

PT New recombinant DNA constructs, for expressing high levels of  
 PT heterologous protein in plasmids of higher plants, includes promoter, a  
 PT leader sequence and a downstream box element -  
 XX  
 PS Disclosure; Fig 33A-B; 164pp; English.

XX  
 CC The present sequence represents a vector of the invention. The  
 CC specification describes recombinant DNA constructs for expressing  
 CC heterologous proteins in the plasmids of higher plants. The DNA  
 CC constructs comprise a 5' regulatory region which includes a promoter  
 CC element, a leader sequence and a downstream box element operably linked  
 CC to a coding region of the heterologous protein. The chimeric regulatory  
 CC region enhances translational efficiency of an mRNA molecule encoded by  
 CC the DNA construct. The DNA constructs are used for producing transformed  
 CC monocot and dicot plants having high levels of heterologous protein  
 CC expression. They can be used to drive expression of proteins with  
 CC agronomic, industrial or pharmaceutical importance, including production  
 CC of vaccines, healthcare products like human hemoglobin, industrial or  
 CC household enzymes. Plants which can be transformed with the constructs  
 CC of the invention include maize, millet, sorghum, sugar cane, rice,  
 CC wheat, barley, oat, rye or turf grass.

XX Sequence 4671 BP; 1278 A; 1181 C; 997 G; 1215 T; 0 other;

Query Match 0.9%; Score 19; DB 21; Length 4671;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1253 aaaaaaatcaatttg 1271  
 ||||||||||||||||  
 DB 2018 AAAAAAAAAATCAATTTC 2000

RESULT 43

AAF22306

ID AAF22306 standard; DNA; 163319 BP.

XX AAF22306;

XX 20-MAR-2001 (first entry)

DE Arabidopsis thaliana chromosome 4 centromere.

XX Centromere; mitochondrion; vector; ds.

OS Arabidopsis thaliana.

XX W020005325-A2.

XX 21-SEP-2000.

PD 17-MAR-2000; 2000WO-US07392.

XX 18-MAR-1999; 99US-0125219.

PR 01-APR-1999; 99US-0127409.

PR 18-MAY-1999; 99US-0134770.

PR 13-SEP-1999; 99US-0153584.

PR 17-SEP-1999; 99US-0154603.

XX (UYCH -) UNIV CHICAGO.

XX  
 PI Preuss D, Copenhagen G, Keith K;  
 XX  
 DR WPI; 2000-587529/55.  
 XX

PT Recombinant DNA construct comprising a plant centromere, useful for  
 PT producing stably inherited mitochondria which can serve as vectors for  
 PT the construction of transgenic plant and animal cells -  
 XX  
 PS Claim 79; Page 1389-1451; 1449pp; English.

XX  
 CC The present invention relates to a recombinant DNA construct of a plant  
 CC (Arabidopsis thaliana) centromere. The constructs are useful for  
 CC producing stably inherited mitochondria which can serve as vectors for  
 CC the construction of transgenic plant and animal cells expressing  
 CC selected proteins such as hormones, enzymes, interleukins, clotting  
 CC factors, cytokines, antibodies, and growth factors.

XX Sequence 163319 BP; 53475 A; 30117 C; 28735 G; 50992 T; 0 other;

Query Match 0.9%; Score 19; DB 21; Length 163319;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1252 taataaaatcaatttt 1270  
 ||||||||||||||||  
 DB 123735 taataaaatcaatttt 123753

RESULT 44

AA12279

ID AA12279 standard; DNA; 126 BP.

XX AA12279;

XX 30-MAR-1999 (first entry)

DE Human diallelic polymorphic DNA fragment WI-18582b.

XX Polymorphism; diallelic; human; forensic; paternity testing; disease;

KW detection; phenotypic typing; characteristic; infection; hereditary;

KW autoimmune disease; cancer; inflammation; drug; therapy; medication;

XX treatment; marker; ss.

OS Homo sapiens.

XX W09820165-A2.

XX 14-MAY-1998.

PD 05-NOV-1997; 97WO-US20313.

XX 06-NOV-1996; 96US-0030455.

XX (WHED ) WHITEHEAD INST BIOMEDICAL RES.

XX Hudson T, Lander ES, Wang D;

XX WPI; 1998-286974/25.

XX New isolated nucleic acid segments from the human genome - used for

XX determining polymorphic forms for use in e.g. forensics, paternity

XX testing or phenotypic typing for disease

XX Claim 1; Page 238; 310pp; English.

XX AAX10269-X12937 are human DNA fragments which contain diallelic

XX polymorphic markers which have been isolated using the primers

XX represented in AAX09121-X10268. The base occupying the polymorphic site

XX is indicated by the appropriate IUPAC-IUB ambiguity code. These fragments

XX can be used in methods for determining polymorphic forms in an individual

XX for use in e.g. forensics, paternity testing or for phenotypic typing for

CC diseases such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan  
 CC syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease,  
 CC familial hypercholesterolemia, polycystic kidney disease, hereditary  
 CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary  
 CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
 CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
 CC autoimmune diseases, inflammatory cancer, diseases of the nervous  
 CC system, infection by pathogenic microorganisms, and characteristics such  
 CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
 CC endurance, fertility, and susceptibility or receptivity to particular  
 CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
 CC segments can also be used to produce medicaments for the treatment of  
 CC prophylaxis of such diseases.

SO Sequence 126 BP; 31 A; 23 C; 20 G; 51 T; 1 other;

Query Match 0.9%; Score 18; DB 19; Length 126;  
 Best Local Similarity 100.0%; Pred. No. 85;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1396 attcattcgtttaatg 1413  
 ||||||||||||||||  
 DB 26 attcattcgtttaatg 43

## RESULT 45

AAAL6184  
 ID AAAL6184 standard; DNA; 622 BP.

XX AAAL6184;

DT 14-JUN-2000 (first entry)

XX Human colon cancer differentially expressed nucleotide sequence #189.

XX Colon cancer; detect; differential expression; human; treatment;

XX detect mutation; non-invasive diagnostic method; ds.

XX Homo sapiens.

XX WO200012702-A2.

PD 09-MAR-2000.

XX 30-AUG-1999; 99WO-US19424.

XX 31-AUG-1998; 98US-0098639.

XX 27-JAN-1999; 99US-0117393.

XX (FARB ) BAYER CORP.

XX Endege WO, Steinmann KE, Astle JH, Burgess CC, Carroll E;

XX Carino TJ, Dwivedi P, Ford DM, Lewis ME, Molino GA, Monahan JE;

XX Schlegel R;

XX WPI: 2000-256641/22.

PS Claim 16; Page 200; 345pp; English.

CC This sequence represents a human nucleotide sequence which is  
 CC differentially expressed in colon cancer cells compared to the expression  
 CC levels in normal cells. The nucleotide sequence can be used as a source  
 CC of primers and probes. The nucleotide sequence is useful for determining  
 CC the phenotype of a cell by detecting the differential expression of the  
 CC sequence relative to a normal cell. The probes derived from the sequence  
 CC can also be used to determine the phenotype of cells in a sample. Probes  
 CC and antibodies which hybridise to the nucleotide sequence can also be  
 CC used to determine the phenotype of a cell. The primers are useful for  
 CC detecting a mutation in a test nucleotide sequence and also for detecting

CC cancer, preferably colon cancer. Antibodies against the protein encoded  
 CC by the nucleotide sequence can also be used in a method to detect colon  
 CC cancer. The diagnostic method is non-invasive and accurate for diagnosing  
 CC colon cancer at an early stage.

SO Sequence 622 BP; 214 A; 103 C; 117 G; 185 T; 3 other;

Query Match 0.9%; Score 18; DB 21; Length 622;  
 Best Local Similarity 100.0%; Pred. No. 83;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1254 aaaaaaalcatttg 1271  
 ||||||||||||||||  
 DB 539 aaaaaaalcatttg 556

## RESULT 46

AA92283/C  
 ID AA92283 standard; DNA; 746 BP.

XX AA92283;

DT 05-JUN-2000 (first entry)

XX Porcine melanocortin-4 receptor gene sequence.

XX Melanocortin-4 receptor; MC4R; pig; fat content; metabolic trait;

XX growth rate; feed consumption; polymorphism identification; lean meat;

XX body weight; ds.

XX Sus scrofa.

XX Key

XX CDS

XX Location/Qualifiers

XX 3..746

XX /tag= a

XX /product= MC4R

XX /note= "Melanocortin-4 receptor. No stop codon given."

XX variation

XX /tag= b

XX WO200006777-A2.

XX 10-FEB-2000.

XX 26-JUL-1999; 99WO-US16862.

XX 27-JUL-1998; 98US-0094287.

XX 15-JAN-1999; 99US-0116186.

XX (IOWA ) UNIV IOWA STATE RES FOUND INC.

XX Rothschild MF, Larson NJ, Kim KS;

XX WPI: 2000-195322/17.

XX P-PSDB; AAY78931, AAY78932.

XX Use of melanocortin-4 receptor gene, for identifying animals,

XX especially pigs, having desired metabolic traits of fat content, growth

XX rate and feed consumption

CC This sequence represents the porcine melanocortin-4 receptor (MC4R) gene.  
 CC MC4R is involved in the mediation of long term weight homeostasis, and  
 CC MC4R antagonists can increase food intake and body weight during chronic  
 CC administration. A polymorphism in the MC4R gene is associated with the  
 CC phenotypic differences in the metabolic traits of fat content, growth  
 CC rate and feed consumption. The pig MC4R gene has a polymorphism at  
 CC position 678 where the nucleotide is G in allele 1 and A in allele 2.  
 CC Allele 1 is associated with leaner animals. The invention relates to a  
 CC method for identifying an animal which possesses a genotype indicative of  
 CC the metabolic traits of fat content, growth rate, and feed consumption.

CC The method comprises obtaining a nucleic acid sample from the animal, and  
 CC identifying a polymorphism in the MCR gene of the sample, which allows  
 CC for the selection of animals with a desired phenotype. The products and  
 CC methods can be used for selecting animals with improved metabolic traits  
 CC of fat content, growth rate, and feed consumption. They are useful in  
 CC breeding programmes to improve fat content, growth rate, and feed  
 CC consumption characteristics in a breed, line or population of animals,  
 CC particularly pigs.

XX  
 XX Sequence 746 BP; 167 A; 196 C; 149 G; 234 T; 0 other;

Query Match 0.98; Score 18; DB 21; Length 746;  
 Best Local Similarity 100.0%; Pred. No. 83;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1934 tataaataaagtaga 1951  
 |||

DB 591 TATAGATATTAGTGA 574

RESULT 47

AA57151  
 ID AAA57151 standard; CDNA: 877 BP.

AC AAA57151;

DT 03-OCT-2000 (first entry)

DE Mouse Psx cDNA.

XX  
 XX Mouse; Psx; homeobox; placenta-specific; placental development;  
 KW gestational trophoblastic disease; cytotlastic; transcriptional regulator;  
 KW abortion induction; gestational trophoblastic tumour;  
 KW invasive mole; choriocarcinoma; placental site trophoblastic tumour; ss.

OS Mus sp.

EH Key Location/Qualifiers

FT CDS 44..787

FT /tag= a

FT /product= "Psx protein"

PN US6063912-A.

PD 16-MAY-2000.

PF 06-AUG-1998; 980S-0129888.

PR 06-AUG-1998; 980S-0129888.

PA (KOKU-) KOREA KUMHO PETROCHEMICAL CO LTD.

PI Chun J, Han Y;

DR WPI: 2000-375541/32.

DR P-PSDB: AAY99886.

XX  
 XX New nucleic acid encoding placenta homeobox protein (psx) useful for  
 PT developing treatment for trophoblast-specific Psx-mediated diseases,  
 PT particularly gestational trophoblastic diseases, e.g. gestational  
 PT trophoblastic tumors

PS Claim 5; Fig 1; 23pp; English.

XX  
 XX The present sequence encodes the mouse Psx protein, a novel  
 CC homeodomain-containing transcription factor. The sequence was isolated  
 CC by screening a 13.5 day-old mouse conceptus cDNA library for  
 CC homeobox-containing sequences using a PCR cloning strategy. Expression  
 CC of the psx polynucleotide is restricted to the placenta, particularly  
 CC to the placenta trophoblast cell layers during embryogenesis. The  
 CC expression pattern of psx may be exploited to detect trophoblast specific  
 CC lineages, such as the labyrinthine trophoblast layer and giant cells. The

CC psx gene and gene product may also be used to develop treatment for  
 CC trophoblast-specific Psx-mediated diseases, such as gestational  
 CC trophoblastic diseases (e.g. gestational trophoblastic tumors, invasive  
 CC moles, choriocarcinomas, and placental site trophoblastic tumors).  
 CC They may also be used to identify putative abortion-inducing agents,  
 CC which may offer an alternative to surgical abortion.

XX  
 XX Sequence 877 BP; 251 A; 204 C; 256 G; 166 T; 0 other;

Query Match 0.98; Score 18; DB 21; Length 877;  
 Best Local Similarity 100.0%; Pred. No. 83;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 861 tggcgggaatgcagtggt 878  
 |||

DB 124 tggcgggaatgcagtggt 141

RESULT 48

AAV7477  
 ID AAV7477 standard; DNA: 1016 BP.

AC AAV7477;

DT 16-MAR-1999 (first entry)

DE Staphylococcus aureus contig SEQ ID #466.

XX  
 XX Computer readable medium; vaccine; S.aureus infection; immunodetection;  
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KW skin infection; surgical wound infection; scalded skin syndrome;  
 KW toxic shock syndrome; ds.

OS Staphylococcus aureus.

EH Key Location/Qualifiers

FT misc\_feature 481..540

FT /tag= a

FT /note= "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"

PN EP786519-A2.

PD 30-JUL-1997.

PF 07-JAN-1997; 97EP-0100117.

PR 05-JAN-1996; 96US-0009861.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;

DR WPI: 1997-374922/35.

XX  
 XX Polynucleotide(s) and proteins derived from Staphylococcus aureus  
 PT stored on computer readable medium and used in the production of  
 PT anti-S.aureus vaccines

PS Claim 1; Page 1382-1383; 3271pp; English.

XX  
 XX This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
 CC of the invention. The DNA sequences are recorded on a computer readable  
 CC medium, preferably selected from a floppy or hard disk, random access  
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 CC the S.aureus DNA sequences allows putative functions to be assigned so  
 CC that protein-encoding or regulatory regions of commercial, therapeutic or  
 CC industrial importance can be obtained. Specifically, sequences which are  
 CC likely to encode antigens have been identified and these polypeptides can

CC be used in a vaccine composition against *S.aureus* infection. The  
CC polypeptides can also be used in a kit for the immunodetection of  
CC *S.aureus* in a sample. *S.aureus* is implicated in numerous human diseases,  
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
CC for recombinant production of the polypeptides. The new DNA sequences  
CC (and their fragments) are useful as primers or probes for isolating  
CC homologues of any of the *S.aureus* DNA sequences contained on the  
CC computer readable medium.

CC Sequence 1016 BP; 334 A; 104 C; 154 G; 359 T; 65 other;

Query Match 0.9%; Score 18; DB 18; Length 1016;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1570 attgtaatgctctaat 1587  
Db 398 attgtaatgctctaat 415  
|||||

RESULT 49

AAC36649  
ID AAC36649 standard; DNA: 1209 BP.

AC AAC36649;

DF 17-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 14573.

XX Hybridisation assay; genetic mapping; gene expression control;  
XX protein identification; signal transduction pathway;  
XX metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

XX EPI033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-030139.

XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0126785.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132485.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142055.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144614.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154039.  
 PR 16-SEP-1999; 99US-0154018.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0156539.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.

PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 0.9%; Score 18; DB 21; Length 1209;  
 Best Local Similarity 100.0%; Pred. No. 83;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1272 tctttatttgtgtgtg 1289  
 |||||||  
 Db 1156 tctttatttgtgtgtg 1173

RESULT 50  
 AAC9977/c  
 ID AAC9977 standard; cDNA; 1274 BP.  
 XX  
 AAC9977;  
 AC 13-MAR-2001 (first entry)  
 XX  
 DE Human secreted protein gene 142 SEQ ID NO:170.  
 XX  
 KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;  
 KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;  
 KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;  
 KW fungicide; ophthalmological; gene therapy; pathological condition;  
 KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;  
 KW neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;  
 KW cerebrovascular disorder; angiogenesis; nervous system disorder;  
 KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
 KW wound healing; skin aging; food additive; preservative;  
 KW Chromosome 20; ss.  
 KW  
 XX Homo sapiens.  
 OS  
 XX MO200070042-A1.  
 PN  
 XX 23-NOV-2000.  
 PD  
 XX 11-MAY-2000; 2000MO-US12788.  
 PE  
 XX 13-MAY-1999; 99US-0134068.  
 PR  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CE;  
 PI Duan RD, Florence KA, Soppet DR;  
 PI WPI; 2000-679828/66.  
 DR P-PSDB; AAB56236.  
 XX  
 PT Isolated nucleic acid molecule encoding a human secreted protein is  
 used in preventing, treating or ameliorating a medical condition -  
 PT  
 XX  
 BS Claim 1; Page 939-940; 1065pp; English.  
 XX  
 CC The polynucleotide sequences given in AAC99818 to AAC99977 encode the  
 CC human secreted proteins given in AAB56077 to AAB56362. Human secreted  
 CC proteins have activities based on the tissues and cells the genes are  
 CC expressed in. Examples of activities include: immunosuppressive;  
 CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;  
 CC vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;  
 CC virucide; fungicide; and ophthalmological. The human secreted  
 CC polynucleotides and proteins can be used to prevent, treat or ameliorate  
 CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,  
 CC dogs, chickens or sheep. They are also used in diagnosing a pathological  
 CC condition or susceptibility to a pathological condition. Disorders which  
 CC are diagnosed or treated include autoimmune diseases e.g. rheumatoid  
 CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or



CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular  
CC disorders e.g. cerebral ischemia, angiogenesis, nervous system disorders  
CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and  
CC fungi and ocular disorders e.g. corneal infection. The proteins can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The proteins can also be used as a  
CC food additive or preservative to increase or decrease storage  
CC capabilities. AAC99809 to AAC99817 and AAB56076 represent sequences used  
CC in the exemplification of the present invention.  
XX

SQ Sequence 1274 BP; 390 A; 256 C; 271 G; 357 T; 0 other;

Query Match 0.9%; Score 18; DB 21; Length 1274;

Best Local Similarity 100.0%; Pred. No. 83;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 ccagagcagtgatcattc 380

|||||

DB 624 CCAGAGCAGTGATTCATTTC 607

Search completed: August 19, 2001, 00:28:55  
Job time: 9281 sec











C 539	14	1.3	2077	US-08-217-327-7	Sequence 7, Appl1	C 612	14	1.3	2688	PCT-US92-02521-3	Sequence 3, Appl1
540	14	1.3	2085	US-08-677-049-1	Sequence 1	613	14	1.3	2698	US-08-677-298-1	Sequence 3, Appl1
541	14	1.3	2095	US-08-405-230-8	Sequence 8, Appl1	614	14	1.3	2743	US-08-525-5568-13	Sequence 13, Appl1
542	14	1.3	2095	US-08-910-990-8	Sequence 9, Appl1	615	14	1.3	2743	US-09-177-8604-13	Sequence 13, Appl1
543	14	1.3	2100	US-08-511-485-9	Sequence 8, Appl1	616	14	1.3	2766	US-08-936-135-3	Sequence 3, Appl1
544	14	1.3	2106	US-09-132-028-1	Sequence 1, Appl1	617	14	1.3	2790	US-08-985-950-21	Sequence 21, Appl1
545	14	1.3	2116	US-08-860-091A-1	Sequence 1, Appl1	618	14	1.3	2813	US-08-344-155C-99	Sequence 99, Appl1
546	14	1.3	2144	US-08-834-306-15	Sequence 15, Appl1	619	14	1.3	2900	US-08-117-362-1	Sequence 1, Appl1
547	14	1.3	2144	US-08-893-674A-15	Sequence 15, Appl1	620	14	1.3	2900	US-08-486-924-1	Sequence 1, Appl1
548	14	1.3	2158	US-07-688-352C-3	Sequence 3, Appl1	621	14	1.3	2912	US-07-931-943-1	Sequence 1, Appl1
549	14	1.3	2158	US-08-474-379C-3	Sequence 3, Appl1	622	14	1.3	2912	US-08-624-601-1	Sequence 4, Appl1
550	14	1.3	2158	US-09-146-249A-3	Sequence 3, Appl1	623	14	1.3	2915	US-07-931-943-4	Sequence 4, Appl1
551	14	1.3	2158	US-08-206-188B-3	Sequence 3, Appl1	624	14	1.3	2965	US-08-460-570-1	Sequence 1, Appl1
552	14	1.3	2158	PCT-US91-02714-3	Sequence 3, Appl1	625	14	1.3	2965	US-08-460-570-2	Sequence 2, Appl1
553	14	1.3	2170	US-09-345-468-11	Sequence 11, Appl1	626	14	1.3	2965	US-08-286-870A-1	Sequence 1, Appl1
554	14	1.3	2205	US-09-345-468-11	Sequence 11, Appl1	627	14	1.3	2965	US-08-286-870A-2	Sequence 2, Appl1
555	14	1.3	2205	PCT-US93-00031-12	Sequence 12, Appl1	628	14	1.3	2965	US-09-110-517-3	Sequence 3, Appl1
556	14	1.3	2208	PCT-US93-00031-14	Sequence 14, Appl1	629	14	1.3	2970	US-09-110-517-3	Sequence 3, Appl1
557	14	1.3	2210	US-08-464-700-53	Sequence 53, Appl1	630	14	1.3	3010	US-08-714-918-71	Sequence 71, Appl1
558	14	1.3	2217	PCT-US93-00031-8	Sequence 8, Appl1	631	14	1.3	3010	US-09-265-315-71	Sequence 71, Appl1
C 559	14	1.3	2220	US-08-891-322-1	Sequence 1, Appl1	C 632	14	1.3	3010	US-09-265-315-71	Sequence 71, Appl1
C 560	14	1.3	2220	US-09-277-019-1	Sequence 1, Appl1	C 633	14	1.3	3010	US-09-265-315-71	Sequence 71, Appl1
561	14	1.3	2230	PCT-US93-00031-16	Sequence 16, Appl1	634	14	1.3	3119	US-09-299-843A-31	Sequence 31, Appl1
562	14	1.3	2230	US-08-217-327-5	Sequence 5, Appl1	635	14	1.3	3119	US-08-781-802-7	Sequence 7, Appl1
563	14	1.3	2236	US-08-484-93B-13	Sequence 13, Appl1	636	14	1.3	3147	US-08-694-078-7	Sequence 7, Appl1
564	14	1.3	2236	US-08-484-158B-13	Sequence 13, Appl1	637	14	1.3	3147	US-09-058-260-7	Sequence 6, Appl1
565	14	1.3	2236	US-08-484-596A-13	Sequence 13, Appl1	C 638	14	1.3	3182	US-08-595-556-6	Sequence 6, Appl1
566	14	1.3	2236	US-08-480-150A-13	Sequence 13, Appl1	639	14	1.3	3327	US-08-372-892-3	Sequence 3, Appl1
567	14	1.3	2236	US-08-458-731-13	Sequence 13, Appl1	640	14	1.3	3354	US-08-372-892-1	Sequence 1, Appl1
568	14	1.3	2242	US-08-149-223A-13	Sequence 13, Appl1	C 641	14	1.3	3367	US-08-372-892-1	Sequence 1, Appl1
569	14	1.3	2247	US-08-641-627A-37	Sequence 37, Appl1	642	14	1.3	3367	PCT-US85-963B-12	Sequence 12, Appl1
570	14	1.3	2254	US-08-852-82A-1	Sequence 27, Appl1	64					







Sequence 10, April
Sequence 130, April
Sequence 39, April
Sequence 65, April
Sequence 66, April
Sequence 67, April
Sequence 68, April
Sequence 9, April
Sequence 9, April
Sequence 9, April
Sequence 4, April
Sequence 65, April
Sequence 66, April
Sequence 67, April
Sequence 68, April
Sequence 9, April
Sequence 4, April
Sequence 32, April
Sequence 33, April
Sequence 34, April
Sequence 35, April
Sequence 13, April
Sequence 14, April
Sequence 13, April

## RESULT 1

ORGANISM: Homo Sapiens  
TISSUE TYPE: Brain  
FEATURE:

```

; IDENTIFICATION METHOD: von Heijne matrix
;
; OTHER INFORMATION: score 12.5
;
; OTHER INFORMATION: seq FVVLALVAGVLG/NE
US-08-905-223-169

```

RESULT 2  
US-09-129-888-1  
; Sequence 1, Application US/09129888B  
; Patent No. 6063912

GENERAL INFORMATION:  
APPLICANT: CHUN, Jong Yoon  
TITLE OF INVENTION: Placenta trophoblast-specific gene  
FILE REFERENCE: 1942/29  
CURRENT APPLICATION NUMBER: US/09/129,888B  
CURRENT FILING DATE: 1998-08-06  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: WordPerfect 6.1 Windows  
SEQ ID NO 1  
LENGTH: 877  
TYPE: DNA  
ORGANISM: mouse  
US-09-129-888-1

Query Match 1.7%; Score 18; DB 3; Length 877;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 765 tggtaggaatgcagtggt 782  
|||||  
DB 124 tggtaggaatgcagtggt 141

RESULT 3  
US-08-574-959A-1  
Sequence 1, Application US/08574959A  
Patent No. 5962224  
GENERAL INFORMATION:  
APPLICANT: Jaekyoon Shin, Insil Jeung, Ratna K. Vadlamudi  
TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street, Suite 510  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/574,959A  
FILING DATE: 19-DEC-95  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: DFN-008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2083 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 67..1390  
US-08-574-959A-1

Query Match 1.7%; Score 18; DB 2; Length 2083;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 143 tgtcatggctctctg 160  
|||||  
DB 1271 tgtcatggctctctg 1288

RESULT 4  
PCT-US93-05640-9/c  
Sequence 9, Application PC/TUS9305640  
GENERAL INFORMATION:  
APPLICANT: University of Connecticut  
TITLE OF INVENTION: Contraceptive Vaccine  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/05640  
FILING DATE: 19930610  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/897,883  
FILING DATE: June 12, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Brook, David E.  
REGISTRATION NUMBER: 22,592  
REFERENCE/DOCKET NUMBER: UCT90-01AA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1104 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
PCT-US93-05640-9

Query Match 1.6%; Score 17; DB 5; Length 1104;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 324 tgttcaaatcattc 340  
|||||  
DB 526 tgttcaaatcattc 510

RESULT 5  
US-08-857-213-2/c  
Sequence 2, Application US/08857213  
Patent No. 6054290  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Murty, Lynn E.  
TITLE OF INVENTION: HUMAN VESICLE BINDING PROTEIN  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA

```

? COUNTRY: USA
? ZIP: 94304
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: DOS
? SOFTWARE: FASTAQ for Windows Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/857,213
? FILING DATE: Herewith
? CLASSIFICATION: 536
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Billings, Lucy J.
? REGISTRATION NUMBER: 36,749
? REFERENCE/DOCKET NUMBER: PF-0297 US
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 415-855-0555
? TELEFAX: 415-845-4166
? TELEX:
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1721 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? IMMEDIATE SOURCE:
? LIBRARY: FIBRNGT01
? CLONE: 148415
? US-08-857-213-2

Query Match      1.6%; Score 17; DB 3; Length 1721;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 683 aaagctgaacattc 699
Db 1599 AAGACTCTGACACTTC 1583

RESULT 6
US-08-942-008-1
? Sequence 1, Application US/08942008
? Patent No. 6133419
? GENERAL INFORMATION:
? APPLICANT: Baselmann, Sylvia
? TITLE OF INVENTION: Nucleotide Sequences that Encode
? TITLE OF INVENTION: Phosphatidylinositol-3' Kinase Associated Proteins and
? NUMBER OF SEQUENCES: 2
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: ONYX Pharmaceuticals, Inc.
? STREET: 3031 Research Drive
? CITY: Richmond
? STATE: CA
? COUNTRY: USA
? ZIP: 94806
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/942,008
? FILING DATE:
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Giotta, Gregory
? REGISTRATION NUMBER: 32,028
? REFERENCE/DOCKET NUMBER: ONYX1027

```

```

? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (510) 262-8710
? TELEFAX: (510) 222-9758
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2307 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? HYPOTHETICAL: NO
? ANTI-SENSE: NO
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 162..1928
? US-08-942-008-1

Query Match      1.6%; Score 17; DB 3; Length 2307;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 676 tatgggaagactcga 692
Db 1458 TATGGGAGACTCTCA 1474

RESULT 7
US-08-670-175-2/c
? Sequence 2, Application US/08670175
? Patent No. 5854081
? GENERAL INFORMATION:
? APPLICANT: LINDEN, JOEL
? APPLICANT: TAYLOR, HEIDI
? APPLICANT: ROBEVA, ANNA
? APPLICANT: WOODARD, ROBIN
? APPLICANT: JIN, XIAOWEI
? TITLE OF INVENTION: STABLE EXPRESSION OF HUMAN ADENOSINE
? TITLE OF INVENTION: RECEPTORS, AND ASSAYS EMPLOYING THE SAME
? NUMBER OF SEQUENCES: 8
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,
? STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
? CITY: ARLINGTON
? STATE: VA
? COUNTRY: USA
? ZIP: 22202
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/670,175
? FILING DATE: 20-JUN-1996
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: OBLON, NORMAN F.
? REGISTRATION NUMBER: 24,618
? REFERENCE/DOCKET NUMBER: 494-176-0
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 703-413-3000
? TELEFAX: 703-413-2220
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 52 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: other nucleic acid
? DESCRIPTION: /desc = "SYNTHETIC DNA PRIMER"
? US-08-670-175-2

```

Query Match 1.5%; Score 16; DB 2; Length 52;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 226 gtcattgctgctgctga 241  
DB 52 gtcattgctgctgctga 37

RESULT 8  
US-08-270-985-4  
; Sequence 4, Application US/08270985  
; Patent No. 5668012  
; GENERAL INFORMATION:  
; APPLICANT: Newman, Peter J.  
; APPLICANT: Gumina, Richard J.  
; APPLICANT: Kirschbaum, Nancy  
; TITLE OF INVENTION: Platelet-Endothelial Cell Adhesion  
; TITLE OF INVENTION: Molecule-1 Compositions and Methods  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: P. O. Box 1497  
; CITY: Madison  
; STATE: Wisconsin  
; COUNTRY: USA  
; ZIP: 53701-1497

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch,  
MEDIUM TYPE: 1.44 MB storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: MS-DOS 3.3  
SOFTWARE: WordPerfect, Version 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/270,985  
FILING DATE: 5-July-94  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Scanlon, William J.  
REGISTRATION NUMBER: 30136  
REFERENCE/DOCKET NUMBER: 30383/137  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (608) 258-4284  
TELEFAX: (608) 258-4258  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 148 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-270-985-4

Query Match 1.5%; Score 16; DB 1; Length 148;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 346 ttattttctggagaa 361  
DB 106 ttattttctggagaa 121

RESULT 9  
US-08-478-208-5  
; Sequence 5, Application US/08478208A  
; Patent No. 6087331  
; GENERAL INFORMATION:  
; APPLICANT: Newman, Peter J.  
; APPLICANT: Gumina, Richard J.

APPLICANT: Kirschbaum, Nancy  
; TITLE OF INVENTION: THERAPEUTIC USE OF PLATELET-ENDOTHELIAL CELL ADHESION  
; FILE REFERENCE: 160180.90147  
; CURRENT APPLICATION NUMBER: US/08/478,208A  
; CURRENT FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 148  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (20)..(127)  
US-08-478-208-5

Query Match 1.5%; Score 16; DB 3; Length 148;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 346 ttattttctggagaa 361  
DB 106 ttattttctggagaa 121

RESULT 10  
US-08-464-164-5/C  
; Sequence 5, Application US/08464164  
; Patent No. 5614195  
; GENERAL INFORMATION:  
; APPLICANT: Tomley, Fiona M.  
; APPLICANT: Dunn, Paul P. J.  
; APPLICANT: Bumstead, Janene M.  
; APPLICANT: Vermeulen, Arno N.  
; TITLE OF INVENTION: Coccidiosis poultry vaccine  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Akzo No. 5614195el Patent Department  
; STREET: 1300 Piccard Drive, Suite 206  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: U.S.A.  
; ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/464,164  
FILING DATE: June 2, 1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Gormley, Mary E.  
REGISTRATION NUMBER: 34,409  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 258-5200  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 250 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Elmeria maxima  
STRAIN: Houghton  
DEVELOPMENTAL STAGE: Sporozoite  
IMMEDIATE SOURCE:

LIBRARY: sporozoitc cDNA cloned in Lambda ZAP11  
CLONE: Em7-0-4, 3' end of clone  
US-08-464-164-5

Query Match 1.5%; Score 16; DB 1; Length 250;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 734 aaagttgcagatga 749  
DB 40 AAAAGTTGCAGATGA 25

## RESULT 11

US-08-338-057-5/c  
Sequence 5, Application US/08338057  
Patent No. 5795741

## GENERAL INFORMATION:

APPLICANT: Tomley, Fiona M.

APPLICANT: Dunn, Paul P. J.

APPLICANT: Bumstead, Janene M.

APPLICANT: Vermeulen, Arno N.

TITLE OF INVENTION: Coccidiosis poultry vaccine

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Organon Teknika Corporation

STREET: 1330 Piccard Drive

CITY: Rockville

STATE: Maryland

COUNTRY: U.S.A.

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/338.057

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: EP 93.309078.9

FILING DATE: 12-NOV-1993

ATTORNEY/AGENT INFORMATION:

NAME: Gormley, Mary E.

REGISTRATION NUMBER: 34,409

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 258-5200

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 250 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Elmeria maxima

STRAIN: Houghton

DEVELOPMENTAL STAGE: Sporozoite

IMMEDIATE SOURCE:

LIBRARY: sporozoite cDNA cloned in Lambda ZAP11

CLONE: Em7-0-4, 3' end of clone

US-08-338-057-5

DB 40 AAAAGTTGCAGATGA 25

## RESULT 12

US-08-668-416-5/c  
Sequence 5, Application US/08668416  
Patent No. 5843722

## GENERAL INFORMATION:

APPLICANT: Tomley, Fiona M.

APPLICANT: Dunn, Paul P. J.

APPLICANT: Bumstead, Janene M.

APPLICANT: Vermeulen, Arno N.

TITLE OF INVENTION: Coccidiosis poultry vaccine

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Akzo No. 5843722el Patent Department

STREET: 1300 Piccard Drive, Suite 206

CITY: Rockville

STATE: Maryland

COUNTRY: U.S.A.

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/668.416

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/464.164

FILING DATE: June 2, 1995

ATTORNEY/AGENT INFORMATION:

NAME: Gormley, Mary E.

REGISTRATION NUMBER: 34,409

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 258-5200

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 250 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Elmeria maxima

STRAIN: Houghton

DEVELOPMENTAL STAGE: Sporozoite

IMMEDIATE SOURCE:

LIBRARY: sporozoite cDNA cloned in Lambda ZAP11

CLONE: Em7-0-4, 3' end of clone

US-08-668-416-5

Query Match 1.5%; Score 16; DB 2; Length 250;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 734 aaagttgcagatga 749  
DB 40 AAAAGTTGCAGATGA 25

## RESULT 13

US-08-514-014-9/c  
Sequence 9, Application US/08514014  
Patent No. 5707829

## GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth

APPLICANT: MCCOY, John  
APPLICANT: Kelleher, Kerry  
APPLICANT: Carlin, McKeough  
TITLE OF INVENTION: DNA SEQUENCES AND SECRETED PROTEINS  
TITLE OF INVENTION: ENCODED THEREBY  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc. -- Legal Affairs  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/514,014  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Brown, Scott A.  
REGISTRATION NUMBER: 32,724  
REFERENCE/DOCKET NUMBER: G16000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8224  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1354 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 75..356  
US-08-514-014-9

Query Match 1.5%; Score 16; DB 1; Length 1354;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 565 gtgctacatgatatt 580  
|||||  
DB 959 gtgctacatgatatt 944

RESULT 14  
US-08-833-823-9/c  
Sequence 9, Application US/08833823  
Patent No. 5968093  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John  
APPLICANT: Kelleher, Kerry  
APPLICANT: Carlin, McKeough  
TITLE OF INVENTION: DNA SEQUENCES AND SECRETED PROTEINS  
TITLE OF INVENTION: ENCODED THEREBY  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc. -- Legal Affairs  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/833,823  
FILING DATE: 10-Apr-1997  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/514,014  
FILING DATE: 11-AUG-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Brown, Scott A.  
REGISTRATION NUMBER: 32,724  
REFERENCE/DOCKET NUMBER: G16000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8224  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1354 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 75..356  
US-08-833-823-9

Query Match 1.5%; Score 16; DB 2; Length 1354;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 565 gtgctacatgatatt 580  
|||||  
DB 959 gtgctacatgatatt 944

RESULT 15  
US-08-557-892-9  
Sequence 9, Application US/08557892  
Patent No. 5968818  
GENERAL INFORMATION:  
APPLICANT: Macphee, Colin Houston  
APPLICANT: Tew, David Graham  
APPLICANT: Southan, Christopher Donald  
APPLICANT: Hickey, Dierdre Mary Bernadette  
APPLICANT: Glozier, Israel Simon  
APPLICANT: Lawrence, Geoffrey Mark Prouse  
APPLICANT: Rice, Simon Quentlyn John  
TITLE OF INVENTION: Compounds  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/557,892  
FILING DATE: 14 No. 5968818ember 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/387,858  
FILING DATE: 14 No. 5968818ember 1994

ATTORNEY/AGENT INFORMATION:  
NAME: Dustman, Wayne J.  
REGISTRATION NUMBER: 33,870  
REFERENCE/DOCKET NUMBER: P30693C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-5023  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1361 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 38..1360  
US-08-557-892-9

Query Match 1.5%; Score 16; DB 2; Length 1361;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 936 gataatgacgcttg 951  
|||||  
DB 302 GATATGATCGCCTTG 317

RESULT 16  
US-08-387-858A-9  
Sequence 9, Application US/08387858A  
Patent No. 5981252  
GENERAL INFORMATION:  
APPLICANT: MacPhee, Colin Houston  
APPLICANT: Tew, David Graham  
APPLICANT: Southan, Christopher Donald  
APPLICANT: Hickey, Dierdre Mary Bernadette  
APPLICANT: Gloger, Israel Simon  
APPLICANT: Lawrence, Geoffrey Mark Prouse  
APPLICANT: Rice, Simon Quentyn John  
TITLE OF INVENTION: Compounds  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/387,858A  
FILING DATE: 24 February 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB94/01374  
FILING DATE: 24 June 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Dustman, Wayne J.  
REGISTRATION NUMBER: 33,870  
REFERENCE/DOCKET NUMBER: P30693  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-5023  
TELEFAX: 610-270-5090  
TELEX:

INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1361 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 38..1360  
US-08-387-858A-9

Query Match 1.5%; Score 16; DB 2; Length 1361;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 936 gataatgacgcttg 951  
|||||  
DB 302 GATATGATCGCCTTG 317

RESULT 17  
US-09-294-384B-9  
Sequence 9, Application US/09294384B  
Patent No. 6177257  
GENERAL INFORMATION:  
APPLICANT: MacPhee, Colin Houston  
APPLICANT: Tew, David Graham  
APPLICANT: Southan, Christopher Donald  
APPLICANT: Hickey, Dierdre Mary Bernadette  
APPLICANT: Gloger, Israel Simon  
APPLICANT: Lawrence, Geoffrey Mark Prouse  
APPLICANT: Rice, Simon Quentyn John  
TITLE OF INVENTION: Compounds  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/294,384B  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/387,858  
FILING DATE: 24 February 1995  
APPLICATION NUMBER: PCT/GB94/01374  
FILING DATE: 24 June 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Dustman, Wayne J.  
REGISTRATION NUMBER: 33,870  
REFERENCE/DOCKET NUMBER: P30693  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-5023  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1361 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear



MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 38..1360  
US-09-294-384B-9

Query Match 1.5%; Score 16; DB 4; Length 1361;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 936 gataatgacgcttg 951  
|||||  
Db 302 GATAATGATGCCCTTG 317

RESULT 18  
US-08-470-187-7  
Sequence 7, Application US/08470187  
Patent No. 5532152

GENERAL INFORMATION:  
APPLICANT: Cousens, Lawrence S.  
APPLICANT: Eberhardt, Christine E.  
APPLICANT: Gray, Patrick W.  
APPLICANT: Tjoelker, Larry W.  
APPLICANT: Wilder, Cheryl L.  
TITLE OF INVENTION: Platelet-Activating Factor Acetyl  
TITLE OF INVENTION: Hydrolase  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gertein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA

ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/470,187  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5532152and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 31672  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: 25-3658

INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1520 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 162..1484  
US-08-470-187-7

Query Match 1.5%; Score 16; DB 1; Length 1520;  
Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 936 gataatgacgcttg 951  
|||||  
Db 426 GATAATGATGCCCTTG 441

RESULT 19  
US-08-318-905-7  
Sequence 7, Application US/08318905  
Patent No. 5641669

GENERAL INFORMATION:  
APPLICANT: Cousens, Lawrence S.  
APPLICANT: Eberhardt, Christine D.  
APPLICANT: Gray, Patrick W.  
APPLICANT: Le Trong, Hai  
APPLICANT: Tjoelker, Larry W.  
APPLICANT: Wilder, Cheryl L.  
TITLE OF INVENTION: Platelet-Activating Factor Acetyl  
TITLE OF INVENTION: Hydrolase  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gertein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA

ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/318,905  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/133,803  
FILING DATE: 6-OCT-1993

ATTORNEY/AGENT INFORMATION:  
NAME: No. 5641669and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32205  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: 25-3658  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1520 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 162..1484  
US-08-318-905-7

Query Match 1.5%; Score 16; DB 1; Length 1520;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 936 gataatgacgcttg 951  
|||||  
Db 426 GATAATGATGCCCTTG 441

RESULT 20  
US-08-483-232-7  
Sequence 7, Application US/08483232

Patent No. 5656431  
GENERAL INFORMATION:  
APPLICANT: Cousens, Lawrence S.  
APPLICANT: Eberhardt, Christine D.  
APPLICANT: Gray, Patrick W.  
APPLICANT: Le Truong, Hai  
APPLICANT: Tjoelker, Larry W.  
TITLE OF INVENTION: Platelet-Activating Factor  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/483,232  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/318,905  
FILING DATE: 06-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/133,803  
FILING DATE: 06-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5656431and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 27866/32689  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: 25-3658  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1520 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 162..1484  
US-08-483-232-7

Query Match 1.5%; Score 16; DB 1; Length 1520;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 936 gataatgacgccttg 951  
|||||  
DB 426 GATATGATCGCCTTG 441

RESULT 21  
US-08-483-140-7  
Sequence 7, Application US/08483140  
Patent No. 5698403  
GENERAL INFORMATION:  
APPLICANT: ICOS Corporation  
TITLE OF INVENTION: Platelet-Activating Factor Acetyl  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/483,140  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/318,905  
FILING DATE: 6-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/133,803  
FILING DATE: 6-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5698403and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32781  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: 25-3658  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1520 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 162..1484  
US-08-483-140-7

Query Match 1.5%; Score 16; DB 1; Length 1520;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 936 gataatgacgccttg 951  
|||||  
DB 426 GATATGATCGCCTTG 441

RESULT 22  
US-08-485-938A-7  
Sequence 7, Application US/08485938A  
Patent No. 5847088  
GENERAL INFORMATION:  
APPLICANT: Cousens, Lawrence S.  
APPLICANT: Eberhardt, Christine D.  
APPLICANT: Gray, Patrick W.  
APPLICANT: Le Truong, Hai  
APPLICANT: Tjoelker, Larry W.  
TITLE OF INVENTION: Platelet-Activating Factor  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,938A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/318,905  
FILING DATE: 06-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/133,803  
FILING DATE: 06-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 584708band, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 27866/32792  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: 25-3658  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1520 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 162..1484  
US-08-938A-7

Query Match 1.5%; Score 16; DB 2; Length 1520;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 936 gataatgacgccttg 951  
|||||  
Db 426 GATAATGATCCCTTG 441

RESULT 23  
US-08-910-041-7  
Sequence 7, Application US/08910041  
Patent No. 5977308  
GENERAL INFORMATION:  
APPLICANT: Cousens, Lawrence S.  
APPLICANT: Eberhardt, Christine D.  
APPLICANT: Gray, Patrick W.  
APPLICANT: Le Trong, Hai  
APPLICANT: Tjoelker, Larry W.  
APPLICANT: Wilder, Cheryl L.  
TITLE OF INVENTION: Platelet-Activating Factor  
TITLE OF INVENTION: Acetylhydrolase  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/910,041  
FILING DATE:

CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/483,232  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/318,905  
FILING DATE: 06-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/133,803  
FILING DATE: 06-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Rin-Laures, Li-Hsien  
REGISTRATION NUMBER: 33,547  
REFERENCE/DOCKET NUMBER: 27866/34026  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: 25-3658  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1520 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 162..1484  
US-08-910-041-7

Query Match 1.5%; Score 16; DB 2; Length 1520;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 936 gataatgacgccttg 951  
|||||  
Db 426 GATAATGATCCCTTG 441

RESULT 24  
US-09-328-474-7  
Sequence 7, Application US/09328474  
Patent No. 6045794  
GENERAL INFORMATION:  
APPLICANT: Cousens, Lawrence S.  
APPLICANT: Eberhardt, Christine D.  
APPLICANT: Gray, Patrick W.  
APPLICANT: Le Trong, Hai  
APPLICANT: Tjoelker, Larry W.  
APPLICANT: Wilder, Cheryl L.  
TITLE OF INVENTION: Platelet-Activating Factor  
TITLE OF INVENTION: Acetylhydrolase  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/328,474  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/483,232  
FILING DATE: 07-JUN-1995

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/318,905  
FILING DATE: 06-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/133,803  
FILING DATE: 06-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Rin-Lauries, Li-Hsien  
REGISTRATION NUMBER: 33,547  
REFERENCE/DOCKET NUMBER: 27866/34026  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: 25-3658  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1520 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 162..1484  
US-09-328-474-7

Query Match 1.5%; Score 16; DB 3; Length 1520;  
Best local similarity 100.0%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 936 gataatgacgccttg 951  
|||||

DB 426 GATAATGATGCGCTTG 441

RESULT 25  
US-09-100-546-7  
Sequence 7, Application US/09100546  
Patent No. 6099836  
GENERAL INFORMATION:  
APPLICANT: Cousens, Lawrence S.  
APPLICANT: Eberhardt, Christine D.  
APPLICANT: Gray, Patrick W.  
APPLICANT: Le Trong, Hai  
APPLICANT: Tjoelker, Larry W.  
APPLICANT: Wilder, Cheryl L.  
TITLE OF INVENTION: Platelet-Activating Factor  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/100,546  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/010,715  
FILING DATE:  
APPLICATION NUMBER: US 08/318,905  
FILING DATE: 06-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/133,803

FILING DATE: 06-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 6099836and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 27866/32793  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: 25-3658  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1520 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 162..1484  
US-09-100-546-7

Query Match 1.5%; Score 16; DB 3; Length 1520;  
Best local similarity 100.0%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 936 gataatgacgccttg 951  
|||||

DB 426 GATAATGATGCGCTTG 441

RESULT 26  
US-09-010-715-7  
Sequence 7, Application US/09010715  
Patent No. 614625  
GENERAL INFORMATION:  
APPLICANT: Cousens, Lawrence S.  
APPLICANT: Eberhardt, Christine D.  
APPLICANT: Gray, Patrick W.  
APPLICANT: Le Trong, Hai  
APPLICANT: Tjoelker, Larry W.  
APPLICANT: Wilder, Cheryl L.  
TITLE OF INVENTION: Platelet-Activating Factor  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/010,715  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/318,905  
FILING DATE: 06-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/133,803  
FILING DATE: 06-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 614625and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 27866/32793  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300

TELEFAX: (312) 474-0448  
TELEX: 25-3658  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1520 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 162..1484  
US-09-010-715-7

Query Match 1.5%; Score 16; DB 3; Length 1520;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 936 gataatgacgcttg 951  
|||||  
DB 426 GATAATGATGCCTTG 441

## RESULT 27

US-09-577-758-7

Sequence 7, Application US/09577758  
Patent No. 6203790  
GENERAL INFORMATION:  
APPLICANT: Cousens, Lawrence S.  
APPLICANT: Eberhardt, Christine D.  
APPLICANT: Gray, Patrick W.  
APPLICANT: Le Trong, Hai  
APPLICANT: Tjoelker, Larry W.  
APPLICANT: Wilder, Cheryl L.  
TITLE OF INVENTION: Platelet-Activating Factor  
TITLE OF INVENTION: Acetylhydrolase  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borum  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/577,758  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/010,715  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/133,803  
FILING DATE: 06-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 6203790and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 27866/32793  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: 25-3658  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1520 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 162..1484  
US-09-577-758-7

Query Match 1.5%; Score 16; DB 4; Length 1520;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 936 gataatgacgcttg 951  
|||||  
DB 426 GATAATGATGCCTTG 441

## RESULT 28

US-08-888-998-1

Sequence 1, Application US/08888998  
Patent No. 6124524  
GENERAL INFORMATION:  
APPLICANT: JAMES, Douglas W.  
APPLICANT: LIM, Eda  
APPLICANT: KELLER, Janis  
APPLICANT: DOONER, Hugo K.  
TITLE OF INVENTION: FAEI GENES AND THEIR USES  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend Kourlie and Crew  
STREET: Stewart Street Tower, One Market Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: US  
ZIP: 94105-1493  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/888,998  
FILING DATE: 07-JUL-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,603  
FILING DATE: 26-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 12176-004300  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1641 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..1641  
OTHER INFORMATION: /product= "FAEI from Arabidopsis"  
US-08-888-998-1

Query Match 1.5%; Score 16; DB 3; Length 1641;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 577 attcaagcttgctgt 592

Db 1139 ATTCAAGCTGCTGT 1154

RESULT 29

US-09-362-633-1  
Sequence 1, Application US/09362633  
Patent No. 618435

GENERAL INFORMATION:  
APPLICANT: JAMES, Douglas W.  
APPLICANT: LIM, Eda  
APPLICANT: KELLER, Janis  
APPLICANT: DOONER, Hugo K.  
TITLE OF INVENTION: PAEL GENES AND THEIR USES  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend Kourie and Crew  
STREET: Stewart Street Tower, One Market Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: US  
ZIP: 94105-1493

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/362,633  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/888,998  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 12176-004300  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1641 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1..1641  
OTHER INFORMATION: /product= "PAEL from Arabidopsis"

US-09-362-633-1  
Query Match 1.5%; Score 16; DB 4; Length 1641;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 577 attcaagctgctgt 592  
Db 1139 ATTCAAGCTGCTGT 1154

RESULT 30  
US-08-822-028-16  
Sequence 16, Application US/08822028  
Patent No. 5993813  
GENERAL INFORMATION:  
APPLICANT: MEZES, PETER S  
APPLICANT: GOURLIE, BRIAN B  
APPLICANT: RIXON, MARK W  
APPLICANT: ANDERSON, WH KERR

APPLICANT: KAPLAN, DONALD A  
APPLICANT: SCHOLIM, JEFFREY  
TITLE OF INVENTION: A NOVEL FAMILY OF HIGH AFFINITY,  
MODIFIED ANTIBODIES FOR CANCER TREATMENT  
NUMBER OF SEQUENCES: 74  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DUANE C ULMER  
STREET: P.O. BOX 1967  
CITY: MIDLAND  
STATE: MICHIGAN  
COUNTRY: USA  
ZIP: 48641-1967  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/822,028  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/040,687  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: ULMER, DUANE C  
REGISTRATION NUMBER: 34,941  
REFERENCE/DOCKET NUMBER: C-37,075C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (517) 636-8104  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1984 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-822-028-16

Query Match 1.5%; Score 16; DB 2; Length 1984;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 607 ctacgaagatcatt 622  
Db 88 CTACGAAGATCATT 103

RESULT 31  
US-08-479-285-16  
Sequence 16, Application US/08479285  
Patent No. 6207815  
GENERAL INFORMATION:  
APPLICANT: MEZES, PETER S  
APPLICANT: GOURLIE, BRIAN B  
APPLICANT: RIXON, MARK W  
APPLICANT: ANDERSON, WH KERR  
APPLICANT: KAPLAN, DONALD A  
APPLICANT: SCHOLIM, JEFFREY  
TITLE OF INVENTION: A NOVEL FAMILY OF HIGH AFFINITY,  
MODIFIED ANTIBODIES FOR CANCER TREATMENT  
NUMBER OF SEQUENCES: 74  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DUANE C ULMER  
STREET: P.O. BOX 1967  
CITY: MIDLAND  
STATE: MICHIGAN  
COUNTRY: USA  
ZIP: 48641-1967  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/479,285  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/040687  
FILING DATE: 31-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: ULMER, DUANE C  
REGISTRATION NUMBER: 34,941  
REFERENCE/DOCKET NUMBER: C-37,075C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (517) 636-8104  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1984 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-479-285-16

Query Match 1.5%; Score 16; DB 4; Length 1984;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 607 ctaccgaagatcatt 622  
|||||  
DB 88 CTACCGACAGATCATT 103

RESULT 32  
US-08-455-073A-5/C  
Sequence 5, Application US/08455073A  
Patent No. 5876949  
GENERAL INFORMATION:  
APPLICANT: Gideon Dreyfuss  
APPLICANT: MIKIKO C. Sloml  
APPLICANT: Yan Zhang  
TITLE OF INVENTION: Fragile X Related Proteins, Compositions And Methods  
TITLE OF INVENTION: Of Making And Using The Same  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5876949rls  
STREET: One Liberty Place, 46th floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Wordperfect 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/455,073A  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Deluca, Mark  
REGISTRATION NUMBER: 33,229  
REFERENCE/DOCKET NUMBER: UPN-2201  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2019 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-455-073A-5

Query Match 1.5%; Score 16; DB 2; Length 2019;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 374 tgcattggctcccg 389  
|||||  
DB 505 TGCAATTGGCTCCCG 490

RESULT 33  
US-07-903-103-1  
Sequence 1, Application US/07903103  
Patent No. 5411860  
GENERAL INFORMATION:  
APPLICANT: VOGELSTEIN, BERT  
APPLICANT: KINZLER, KENNETH  
TITLE OF INVENTION: AMPLIFICATION OF HUMAN MDM2 GENE IN  
TITLE OF INVENTION: HUMAN TUMORS  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BANNER, BIRCH, MCKIE AND BECKETT  
STREET: 1001 G ST., N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4597  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/903,103  
FILING DATE: 19920623  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/867,840  
FILING DATE: 07-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: KAGAN, SARAH A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 01107,40148  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
TELEX: 197430 BBMB UT  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2372 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
CELL LINE: Caco-2  
POSITION IN GENOME:  
MAP POSITION: 12q12-14  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 312..1784  
US-07-903-103-1

Query Match 1.5%; Score 16; DB 1; Length 2372;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 152 gctctctgtgaaga 167

Db 580 GCTTCTGTGAAGA 595

## RESULT 34

```

US-08-044-619A-1
; Sequence 1, Application US/08044619A
; Patent No. 5420263
; GENERAL INFORMATION:
; APPLICANT: THE JOHNS HOPKINS UNIVERSITY
; APPLICANT: 720 RUTLAND AVENUE, BALTIMORE, MARYLAND 21205 USA
; TITLE OF INVENTION: AMPLIFICATION OF HUMAN MDM2 GENE IN
; TITLE OF INVENTION: HUMAN TUMORS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BANNER, BIRCH, MCKIE AND BECKETT
; STREET: 1001 G ST., N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4597
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/044,619A
; FILING DATE: 07-APR-1993
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/903,103
; FILING DATE: 23-JUN-1992
; APPLICATION NUMBER: US 07/867,840
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KAGAN, SARAH A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 01107,40148
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; TELEX: 197430 BBMB UT
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2372 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; CELL LINE: Caco-2
; POSITION IN GENOME:
; MAP POSITION: 12q12-14
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 312..1784
; US-08-044-619A-1

```

Query Match 1.5%; Score 16; DB 1; Length 2372;  
 Best Local Similarity 100.0%; Pred. No. 99;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 152 gcttctgtgaaga 167  
 ||||||||||||||||  
 Db 580 GCTTCTGTGAAGA 595

## RESULT 35

```

US-08-283-911-1
; Sequence 1, Application US/08283911
; Patent No. 5519118
; GENERAL INFORMATION:
; APPLICANT: VOGELSTEIN, BERT
; APPLICANT: KINZLER, KENNETH
; TITLE OF INVENTION: AMPLIFICATION OF HUMAN MDM2 GENE IN
; TITLE OF INVENTION: HUMAN TUMORS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BANNER, BIRCH, MCKIE AND BECKETT
; STREET: 1001 G ST., N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4597
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/283,911
; FILING DATE:
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/903,103
; FILING DATE: 23-JUN-1992
; APPLICATION NUMBER: US 07/867,840
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KAGAN, SARAH A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 01107,40148
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; TELEX: 197430 BBMB UT
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2372 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; CELL LINE: Caco-2
; POSITION IN GENOME:
; MAP POSITION: 12q12-14
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 312..1784
; US-08-283-911-1

```

Query Match 1.5%; Score 16; DB 1; Length 2372;  
 Best Local Similarity 100.0%; Pred. No. 99;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 152 gcttctgtgaaga 167  
 ||||||||||||||||  
 Db 580 GCTTCTGTGAAGA 595

RESULT 36  
 US-08-245-500A-2  
 ; Sequence 2, Application US/08245500A  
 ; Patent No. 5550023  
 ; GENERAL INFORMATION:  
 ; APPLICANT: BURRELL, MARILEE



APPLICANT: HILL, DAVID E.  
APPLICANT: KINZLER, KENNETH W.  
APPLICANT: VOGELSTEIN, BERT  
TITLE OF INVENTION: AMPLIFICATION OF HUMAN MDM2 GENE IN  
HUMAN TUMORS  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BANNER, BIRCH, MCKIE AND BECKETT  
STREET: 1001 G STREET, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/245,500A  
FILING DATE: 07-APR-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: KAGAN, SARAH A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 01107,42798  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
TELEX: 197430 BBMB UT  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2372 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
CELL LINE: Caco-2  
POSITION IN GENOME:  
MAP POSITION: 12q12-14  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 312..1784  
US-08-245-500A-2

Query Match 1.5%; Score 16; DB 1; Length 2372;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 gctctctgtgaaga 167  
|||||  
DB 580 gcttctctgtgaaga 595

RESULT 37  
US-08-390-546-2  
Sequence 2, Application US/08390546  
Patent No. 5606044  
GENERAL INFORMATION:  
APPLICANT: BURRELL, MARILEE  
APPLICANT: HILL, DAVID E.  
APPLICANT: KINZLER, KENNETH W.  
APPLICANT: VOGELSTEIN, BERT  
TITLE OF INVENTION: AMPLIFICATION OF HUMAN MDM2 GENE IN  
HUMAN TUMORS  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BANNER, BIRCH, MCKIE AND BECKETT

STREET: 1001 G STREET, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/390,546  
FILING DATE: 07-APR-1993  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: KAGAN, SARAH A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 01107,42798  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
TELEX: 197430 BBMB UT  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2372 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
CELL LINE: Caco-2  
POSITION IN GENOME:  
MAP POSITION: 12q12-14  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 312..1784  
US-08-390-546-2

Query Match 1.5%; Score 16; DB 1; Length 2372;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 gctctctgtgaaga 167  
|||||  
DB 580 gcttctctgtgaaga 595

RESULT 38  
US-08-390-479A-2  
Sequence 2, Application US/08390479A  
Patent No. 5618921  
GENERAL INFORMATION:  
APPLICANT: BURRELL, MARILEE  
APPLICANT: HILL, DAVID E.  
APPLICANT: KINZLER, KENNETH W.  
APPLICANT: VOGELSTEIN, BERT  
TITLE OF INVENTION: AMPLIFICATION OF HUMAN MDM2 GENE IN  
HUMAN TUMORS  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BANNER & MITCOFF, LTD.  
STREET: 1001 G STREET, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/390,479A  
FILING DATE: 02-FEB-1995  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: KAGAN, SARAH A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 01107,48992  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
TELEX: 197430 BRMB UR  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2372 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
CELL LINE: CaCo-2  
POSITION IN GENOME:  
MAP POSITION: 12q12-14  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 312..1784  
US-08-390-479A-2

Query Match 1.5%; Score 16; DB 1; Length 2372;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 gcttcctctgaaaga 167  
|||||  
DB 580 GCTTCTCTGTGAAGA 595

RESULT 39  
US-08-557-393-2  
Sequence 2, Application US/08557393  
Patent No. 5702903  
GENERAL INFORMATION:  
APPLICANT: BURRELL, MARILEE  
APPLICANT: HILL, DAVID E.  
APPLICANT: KINZLER, KENNETH W.  
APPLICANT: VOGELSTEIN, BERT  
TITLE OF INVENTION: AMPLIFICATION OF HUMAN MDM2 GENE IN  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BANNER, BIRCH, MCKIE AND BECKETT  
STREET: 1001 G STREET, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/557,393  
FILING DATE: 13-NOV-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/245,500

FILING DATE: 18-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: KAGAN, SARAH A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 01107,42798  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
TELEX: 197430 BRMB UR  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2372 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
CELL LINE: CaCo-2  
POSITION IN GENOME:  
MAP POSITION: 12q12-14  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 312..1784  
US-08-557-393-2

Query Match 1.5%; Score 16; DB 1; Length 2372;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 gcttcctctgaaaga 167  
|||||  
DB 580 GCTTCTCTGTGAAGA 595

RESULT 40  
US-08-390-516C-2  
Sequence 2, Application US/08390516C  
Patent No. 5708136  
GENERAL INFORMATION:  
APPLICANT: BURRELL, MARILEE  
APPLICANT: HILL, DAVID E.  
APPLICANT: KINZLER, KENNETH W.  
APPLICANT: VOGELSTEIN, BERT  
TITLE OF INVENTION: AMPLIFICATION OF HUMAN MDM2 GENE IN  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BANNER, BIRCH, MCKIE AND BECKETT  
STREET: 1001 G STREET, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/390,516C  
FILING DATE: 07-APR-1993  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: KAGAN, SARAH A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 01107,42798  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299

TELEX: 197430 BBMB UT  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2372 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
CELL LINE: Caco-2  
POSITION IN GENOME:  
MAP POSITION: 12q12-14  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 312..1784  
US-08-390-516C-2

Query Match 1.5%; Score 16; DB 1; Length 2372;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 152 gctctctgtgaaga 167  
|||||  
Db 580 GCTTCTCTGTGAAGA 595

RESULT 41  
US-08-390-517A-2  
Sequence 2, Application US/08390517A  
Patent No. 5736338  
GENERAL INFORMATION:  
APPLICANT: BURRELL, MARILEE  
APPLICANT: HILL, DAVID E.  
APPLICANT: KINZLER, KENNETH W.  
APPLICANT: VOGELSTEIN, BERT  
TITLE OF INVENTION: AMPLIFICATION OF HUMAN MDM2 GENE IN  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BANNER, BIRCH, MCKIE AND BECKETT  
STREET: 1001 G STREET, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/390.517A  
FILING DATE: 07-APR-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: KAGAN, SARAH A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 01107.42798  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
TELEX: 197430 BBMB UT  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2372 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA

HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
CELL LINE: Caco-2  
POSITION IN GENOME:  
MAP POSITION: 12q12-14  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 312..1784  
US-08-390-517A-2  
Query Match 1.5%; Score 16; DB 1; Length 2372;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 152 gctctctgtgaaga 167  
|||||  
Db 580 GCTTCTCTGTGAAGA 595

RESULT 42  
US-08-390-515A-2  
Sequence 2, Application US/08390515A  
Patent No. 5756455  
GENERAL INFORMATION:  
APPLICANT: BURRELL, MARILEE  
APPLICANT: HILL, DAVID E.  
APPLICANT: KINZLER, KENNETH W.  
APPLICANT: VOGELSTEIN, BERT  
TITLE OF INVENTION: AMPLIFICATION OF HUMAN MDM2 GENE IN  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BANNER, BIRCH, MCKIE AND BECKETT  
STREET: 1001 G STREET, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/390.515A  
FILING DATE: 07-APR-1993  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: KAGAN, SARAH A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 01107.42798  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
TELEX: 197430 BBMB UT  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2372 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
CELL LINE: Caco-2  
POSITION IN GENOME:  
MAP POSITION: 12q12-14  
FEATURE:

NAME/KEY: CDS  
LOCATION: 312..1784  
US-08-390-515A-2

Query Match 1.5%; Score 16; DB 1; Length 2372;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 gcttctctgtgaaga 167  
|||||  
Db 580 GCTTCTCTGTGAAGA 595

## RESULT 43

US-08-801-718-2  
Sequence 2, Application US/08801718  
Patent No. 5858976  
GENERAL INFORMATION:  
APPLICANT: BURRELL, MARILEE  
APPLICANT: HILL, DAVID E.  
APPLICANT: KINZLER, KENNETH W.  
APPLICANT: VOGELSTEIN, BERT  
TITLE OF INVENTION: AMPLIFICATION OF HUMAN MDM2 GENE IN  
TITLE OF INVENTION: HUMAN TUMORS  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BANNER, BIRCH, MCKIE AND BECKETT  
STREET: 1001 G STREET, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/801,718  
FILING DATE: 14-FEB-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/390,515  
FILING DATE: 07-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: KAGAN, SARAH A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 01107,42798  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
TELEX: 197430 BBMB UT  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2372 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
CELL LINE: Caco-2  
POSITION IN GENOME:  
MAP POSITION: 12q12-14  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 312..1784  
US-08-801-718-2

Query Match 1.5%; Score 16; DB 2; Length 2372;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 gcttctctgtgaaga 167  
|||||  
Db 580 GCTTCTCTGTGAAGA 595

## RESULT 44

US-09-073-567-1  
Sequence 1, Application US/09073567  
Patent No. 6013786  
GENERAL INFORMATION:  
APPLICANT: Jiandong Chen  
APPLICANT: Sudhir Agrawal  
APPLICANT: Ruiwen Zhang  
TITLE OF INVENTION: MDM2-SPECIFIC ANTISENSE OLIGONUCLEOTIDES  
NUMBER OF SEQUENCES: 49  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff  
STREET: 300 South Wacker Drive, 32nd Floor  
CITY: Chicago  
STATE: IL  
COUNTRY: United States of America  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Microsoft Word 97  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/073,567  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Greenfield, Michael S.  
REGISTRATION NUMBER: 37,147  
REFERENCE/DOCKET NUMBER: 98,057-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 913-0001  
TELEFAX: (312) 913-0002  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2372 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: hmdm2 DNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-09-073-567-1

Query Match 1.5%; Score 16; DB 3; Length 2372;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 gcttctctgtgaaga 167  
|||||  
Db 580 GCTTCTCTGTGAAGA 595

RESULT 45  
US-09-280-805-1  
Sequence 1, Application US/09280805  
Patent No. 6184212  
GENERAL INFORMATION:  
APPLICANT: Loren J. Miraglia, Pamela Nero, Mark J.  
APPLICANT: Graham, Brett P. Monia  
TITLE OF INVENTION: ANTISENSE MODULATION OF HUMAN MDM2  
TITLE OF INVENTION: EXPRESSION  
NUMBER OF SEQUENCES: 271

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Law Offices of Jane Massey Licata  
;; STREET: 66 East Main Street  
;; CITY: Marlton  
;; STATE: NJ  
;; COUNTRY: U.S.A.  
;; ZIP: 08053  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
;;  
;; COMPUTER: IBM PC  
;; OPERATING SYSTEM: WINDOWS 95  
;; SOFTWARE: WORDPERFECT 6.0  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/280,805  
;; FILING DATE: herewith  
;; CLASSIFICATION:  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: 09/048,810  
;; FILING DATE: March 26, 1998  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Licata, Jane Massey  
;; REGISTRATION NUMBER: 32,257  
;; REFERENCE/DOCKET NUMBER: ISPH-0346  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 609-810-1515  
;; TELEFAX: 609-810-1454  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2372 base pairs  
;; TYPE: Nucleic Acid  
;; STRANDEDNESS: Single  
;; TOPOLOGY: Unknown  
;; ANTI-SENSE: NO  
;; PUBLICATION INFORMATION:  
;; AUTHORS: Oliner, J.D.  
;; AUTHORS: Kinzler, K.W.  
;; AUTHORS: Meltzer, P.S.  
;; AUTHORS: George, D.L.  
;; AUTHORS: Vogelstein, B.  
;; TITLE: Amplification of a gene encoding a  
;; TITLE: p53-associated protein in human sarcomas  
;; JOURNAL: Nature  
;; VOLUME: 358  
;; ISSUE: 6381  
;; PAGES: 80-83  
;; DATE: 02-JUL-1992  
;;  
US-09-280-805-1  
;;  
Query Match 1.5%; Score 16; DB 4; Length 2372;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 152 gctctctgtgaaga 167  
DB 580 GCTTCTCTGTGAAGA 595  
;;  
RESULT 46  
US-09-048-810-1  
;; Sequence 1, Application US/09048810  
;; Patent No. 6238921  
;; GENERAL INFORMATION:  
;; APPLICANT: Loren J. Miraglia, Pamela Nero, Mark J.  
;; APPLICANT: Graham, Brett P. Monia  
;; TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDE  
;; TITLE OF INVENTION: MODULATION OF HUMAN MDX2 EXPRESSION  
;; NUMBER OF SEQUENCES: 32  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Law Offices of Jane Massey Licata  
;; STREET: 66 East Main Street  
;; CITY: Marlton  
;; STATE: NJ

;; COUNTRY: U.S.A.  
;; ZIP: 08053  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
;;  
;; COMPUTER: IBM 486  
;; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS  
;; SOFTWARE: WORDPERFECT 5.1  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/048,810  
;; FILING DATE: herewith  
;; CLASSIFICATION:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Licata, Jane Massey  
;; REGISTRATION NUMBER: 32,257  
;; REFERENCE/DOCKET NUMBER: ISPH-0302  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 609-779-2400  
;; TELEFAX: 609-810-1454  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2372 base pairs  
;; TYPE: Nucleic Acid  
;; STRANDEDNESS: Single  
;; TOPOLOGY: Unknown  
;; ANTI-SENSE: NO  
;; PUBLICATION INFORMATION:  
;; AUTHORS: Oliner, J.D.  
;; AUTHORS: Kinzler, K.W.  
;; AUTHORS: Meltzer, P.S.  
;; AUTHORS: George, D.L.  
;; AUTHORS: Vogelstein, B.  
;; TITLE: Amplification of a gene encoding a  
;; TITLE: p53-associated protein in human sarcomas  
;; JOURNAL: Nature  
;; VOLUME: 358  
;; ISSUE: 6381  
;; PAGES: 80-83  
;; DATE: 02-JUL-1992  
;;  
US-09-048-810-1  
;;  
Query Match 1.5%; Score 16; DB 4; Length 2372;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 152 gctctctgtgaaga 167  
DB 580 GCTTCTCTGTGAAGA 595  
;;  
RESULT 47  
US-09-044-506A-1  
;; Sequence 1, Application US/09044506A  
;; Patent No. 5955443  
;; GENERAL INFORMATION:  
;; APPLICANT: C. Frank Bennett, Thomas Condon,  
;; APPLICANT: Shin Flournoy, Hong Zhang  
;; TITLE OF INVENTION: Antisense Modulation of PECAM-1  
;; NUMBER OF SEQUENCES: 68  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Jane Massey Licata, Esq.  
;; STREET: 66 E. Main Street  
;; CITY: Marlton  
;; STATE: NJ  
;; COUNTRY: USA  
;; ZIP: 08053  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
;;  
;; COMPUTER: IBM 486  
;; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS  
;; SOFTWARE: WORDPERFECT 5.1  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/044,506A

FILED DATE: March 19, 1998  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Jane Massey Licata  
REGISTRATION NUMBER: 32,257  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (609) 779-2400  
TELEFAX: (609) 810-1454  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2557 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Unknown  
MOLECULE TYPE: mRNA  
ANTI-SENSE: NO  
PUBLICATION INFORMATION:  
AUTHORS: Newman, P. J.  
AUTHORS: Berndt, M. C.  
AUTHORS: Gorski, J.  
AUTHORS: White, G. C. II.  
AUTHORS: Lyman, S.  
AUTHORS: Paddock, C.  
AUTHORS: Muller, W. A.  
TITLE: pECAM-1 (CD31) Cloning and relation to  
TITLE: adhesion molecules of the immunoglobulin gene  
TITLE: superfamily  
JOURNAL: Science  
VOLUME: 247  
ISSUE: 4947  
PAGES: 1219-1222  
DATE: 09-MAR-1990  
US-09-044-506A-1

Query Match 1.5%; Score 16; DB 2; Length 2557;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 346 ttatttctgagaa 361  
|||||  
Db 2007 TTATTCTGAGAA 2022

RESULT 48  
US-08-478-208-31  
Sequence 31, Application US/08478208A  
Patent No. 6087331  
GENERAL INFORMATION:  
APPLICANT: Newman, Peter J.  
APPLICANT: Gumina, Richard J.  
APPLICANT: Kirschbaum, Nancy  
TITLE OF INVENTION: THERAPEUTIC USE OF PLATELET-ENDOTHELIAL CELL ADHESION  
TITLE OF INVENTION: MOLECULE-1 COMPOSITIONS  
FILE REFERENCE: 160180.90147  
CURRENT APPLICATION NUMBER: US/08/478,208A  
CURRENT FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 31  
LENGTH: 2557  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (142)..(2358)  
US-08-478-208-31

Query Match 1.5%; Score 16; DB 3; Length 2557;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 346 ttatttctgagaa 361  
|||||  
Db 2007 ttatttctgagaa 2022

RESULT 49  
5264554-1  
Patent No. 5264554  
APPLICANT: Newman, Peter J.  
TITLE OF INVENTION: PLATELET CELL ADHESION MOLECULE AND  
VARIANTS THEREOF  
NUMBER OF SEQUENCES: 2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/466,140  
FILING DATE: 19-JAN-1990  
SEQ ID NO:1:  
LENGTH: 2557  
5264554-1

Query Match 1.5%; Score 16; DB 6; Length 2557;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 346 ttatttctgagaa 361  
|||||  
Db 2007 ttatttctgagaa 2022

RESULT 50  
US-08-437-607A-3/C  
Sequence 3, Application US/08437607A  
Patent No. 595579  
GENERAL INFORMATION:  
APPLICANT: Leonard, James N. Montminy, Marc R.  
TITLE OF INVENTION: ISLET-SPECIFIC HOMOPROTEIN AND TRANSCRIPTIONAL  
TITLE OF INVENTION: REGULATOR OF INSULIN GENE EXPRESSION  
NUMBER OF SEQUENCES: 49  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauder & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/437,607A  
FILING DATE: MAY 9, 1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 110-1-001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2917 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)

```

: DESCRIPTION: Human Hoxb13 gene
:
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 582..1184
: FEATURE:
: NAME/KEY: Intron
: LOCATION: 1185..2132
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 2133..2384
:
US-08-437-607A-3

```

```

Query Match      1.5%; Score 16; DB 2; Length 2917;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 414 gaagcgaactcactg 429
   |||||||
Db 968 GAAGCAACTCAGTG 953

```

Search completed: August 19, 2001, 00:19:56  
 Job time: 8917 sec







C 85	16	1.5	669	19	AAV04221	DNA encoding modif
C 86	16	1.5	681	16	AAO87262	Human double mint
C 87	16	1.5	681	16	AAO92516	Human double mint
C 88	16	1.5	719	21	AAE13398	Aspergillus oryzae
C 89	16	1.5	732	21	AAA02380	Human colon cancer
C 90	16	1.5	812	21	AAZ95008	CSG Pro113 EST use
C 91	16	1.5	852	16	AAO87261	Human double mint
C 92	16	1.5	852	16	AAO92515	Human double mint
C 93	16	1.5	888	20	AAH86436	DNA encoding a tum
C 94	16	1.5	949	21	AAAC1922	Arabidopsis thalia
C 95	16	1.5	959	21	AAAC1922	Arabidopsis thalia
C 96	16	1.5	1029	21	AAAC1922	Arabidopsis thalia
C 97	16	1.5	1029	21	AAAC1922	Arabidopsis thalia
C 98	16	1.5	1029	21	AAAC1922	Arabidopsis thalia
C 99	16	1.5	1029	21	AAAC1922	Arabidopsis thalia
C 100	16	1.5	1193	21	AAAC1922	Arabidopsis thalia
C 101	16	1.5	1228	21	AAAC1922	Arabidopsis thalia
C 102	16	1.5	1256	14	AAO38083	Arabidopsis thalia
C 103	16	1.5	1256	14	AAO38083	Arabidopsis thalia
C 104	16	1.5	1302	16	AAO87263	Kangaroo mu-crysta
C 105	16	1.5	1302	16	AAO92517	Human double mint
C 106	16	1.5	1326	21	AAO92517	Human double mint
C 107	16	1.5	1354	18	AAO87431	Arabidopsis thalia
C 108	16	1.5	1354	18	AAO87431	Arabidopsis thalia
C 109	16	1.5	1361	16	AAO81780	Human non-FLR CXG
C 110	16	1.5	1361	16	AAO81780	Human T-cell lymph
C 111	16	1.5	1371	20	AAAC3332	DNA encoding huma
C 112	16	1.5	1419	22	AAE63742	Human CXG chemok
C 113	16	1.5	1476	18	AAE63742	Drosophila gustato
C 114	16	1.5	1476	18	AAE63742	Murine double min
C 115	16	1.5	1486	22	AAO84596	Human KMT2 protei
C 116	16	1.5	1520	16	AAO87947	Human zslg6 DNA
C 117	16	1.5	1520	18	AAO87948	Human platelet act
C 118	16	1.5	1520	18	AAO87948	Human platelet act
C 119	16	1.5	1520	18	AAO87948	Human platelet act
C 120	16	1.5	1520	19	AAO87948	Human platelet act
C 121	16	1.5	1520	20	AAO87948	Human platelet act
C 122	16	1.5	1520	20	AAO87948	Human platelet act
C 123	16	1.5	1520	21	AAO87948	Human platelet act
C 124	16	1.5	1520	21	AAO87948	Human platelet act
C 125	16	1.5	1520	21	AAO87948	Human platelet act
C 126	16	1.5	1520	21	AAO87948	Human platelet act
C 127	16	1.5	1520	21	AAO87948	Human platelet act
C 128	16	1.5	1520	21	AAO87948	Human platelet act
C 129	16	1.5	1520	21	AAO87948	Human platelet act
C 130	16	1.5	1520	21	AAO87948	Human platelet act
C 131	16	1.5	1520	21	AAO87948	Human platelet act
C 132	16	1.5	1520	21	AAO87948	Human platelet act
C 133	16	1.5	1520	21	AAO87948	Human platelet act
C 134	16	1.5	1520	21	AAO87948	Human platelet act
C 135	16	1.5	1520	21	AAO87948	Human platelet act
C 136	16	1.5	1520	21	AAO87948	Human platelet act
C 137	16	1.5	1520	21	AAO87948	Human platelet act
C 138	16	1.5	1520	21	AAO87948	Human platelet act
C 139	16	1.5	1520	21	AAO87948	Human platelet act
C 140	16	1.5	1520	21	AAO87948	Human platelet act
C 141	16	1.5	1520	21	AAO87948	Human platelet act
C 142	16	1.5	1520	21	AAO87948	Human platelet act
C 143	16	1.5	1520	21	AAO87948	Human platelet act
C 144	16	1.5	1520	21	AAO87948	Human platelet act
C 145	16	1.5	1520	21	AAO87948	Human platelet act
C 146	16	1.5	1520	21	AAO87948	Human platelet act
C 147	16	1.5	1520	21	AAO87948	Human platelet act
C 148	16	1.5	1520	21	AAO87948	Human platelet act
C 149	16	1.5	1520	21	AAO87948	Human platelet act
C 150	16	1.5	1520	21	AAO87948	Human platelet act
C 151	16	1.5	1520	21	AAO87948	Human platelet act
C 152	16	1.5	1520	21	AAO87948	Human platelet act
C 153	16	1.5	1520	21	AAO87948	Human platelet act
C 154	16	1.5	1520	21	AAO87948	Human platelet act
C 155	16	1.5	1520	21	AAO87948	Human platelet act
C 156	16	1.5	1520	21	AAO87948	Human platelet act
C 157	16	1.5	1520	21	AAO87948	Human platelet act

















OY 961 gtgattatcactcttaacatatttggaacatgagcttgatagatagatcattat 1020  
 Db 1057 gtgattatcactcttaacatatttggaacatgagcttgatagatagatcattat 1116  
 OY 1021 aggatgacaacacgaagattcgaatgat 1050  
 Db 1117 aggatgacaacacgaagattcgaatgat 1146

RESULT 3

AAF97893  
 ID AAF97893 standard; cDNA; 2108 BP.

AC AAF97893;

DT 01-JUN-2001 (first entry)

XX Human secreted protein cDNA, SEQ ID NO: 20.

XX Human: secreted protein; immunomodulatory; antisclerotic;  
 XX dermatological; anti-inflammatory; anti-HIV; cytostatic; cardiant;  
 XX vascular; anti-angiogenic; ophthalmological; neuroprotectant;  
 XX neurotropic; anticonvulsant; antialzheimers; antiparkinsonian;  
 XX antimicrobial; vulnerary; vaccine; gene therapy; cancer;  
 XX protein coordinate data; infection; ss.

OS Homo sapiens.

PN MO200121658-A1.

PD 29-MAR-2001.

PF 22-SEP-2000; 2000MO-US26013.

PR 24-SEP-1999; 99US-0155709.

PI (HUMA-) HUMAN GENOME SCI INC.

PI Ni J, Baker KP, Birse CE, Ebner R, Fiscella M, Komatsoulis GA;  
 PI Lafleur DM, Moore PA, Olsen HS, Rosen CA, Ruben SA, Soppet DR;  
 PI Young PE, Wei P, Florence KA;

DR WPI; 2001-235311/24.

PT Nucleic acids encoding 32 human secreted polypeptides, useful for  
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's  
 PT disease and diabetic retinopathy -

PS Claim 1; Page 726-727; 890pp; English.

XX The present sequence encodes one of 32 novel human secreted polypeptides.  
 CC The nucleic acid molecules and polypeptides they encode may be used in  
 CC the prevention, diagnosis and treatment of diseases such as  
 CC immune disorders (e.g. multiple sclerosis, systemic lupus erythematosus  
 CC and human immuno-deficiency virus (HIV) infections), hyperproliferative  
 CC disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases  
 CC (e.g. Schmitt's syndrome, Chaga's cardiomyopathy and coronary  
 CC arteriosclerosis), angiogenic disorders (e.g. corneal graft  
 CC neovascularisation and diabetic retinopathy), neurological disorders  
 CC (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),  
 CC infectious diseases and/or for promoting wound healing, regeneration  
 CC and/or chemotaxis. The nucleic acid molecules may be used to produce the  
 CC secreted polypeptides. They may also be used as DNA probes in diagnostic  
 CC assays to detect and quantitate the presence of similar nucleic acid  
 CC sequences in samples. The polypeptides may be used as antigens in the  
 CC production of antibodies and in assays to identify modulators of  
 CC their expression and activity.

SQ Sequence 2108 BP; 592 A; 389 C; 460 G; 664 T; 3 other;

Query Match

100.0%; Score 1050; DB 22; Length 2108;

Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1050; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atgacgtgttctgctgctcctgagcttggctggcggtgttttgggaaagagattact 60  
 Db 130 atgacgtgttctgctgctcctgagcttggctggcggtgttttgggaaagagattact 189  
 OY 61 atattaaatacacaaggtctgttcttcggaatggaattggcctataccagaagag 120  
 Db 190 atattaaatacacaaggtctgttcttcggaatggaattggcctataccagaagag 249  
 OY 121 cggatccagaagctgctgctcctgagcttgcagagctctcgtgaagaagagccttttgg 180  
 Db 250 cggatccagaagctgctgctcctgagcttgcagagctctcgtgaagaagagccttttgg 309  
 OY 181 ccagagctcgagtggtggttaactgtttcactcgtcctcgggtcaccgtatgtatgtg 240  
 Db 310 ccagagctcgagtggtggttaactgtttcactcgtcctcgggtcaccgtatgtatgtg 369  
 OY 241 aagggaatgacaacactggctcaccacagcagtgatcttgcacaccttggagagat 300  
 Db 370 aagggaatgacaacactggctcaccacagcagtgatcttgcacaccttggagagat 429  
 OY 301 ggaatcctttttagcttgcagagtggttgcaaatcattcattcactccttattcttgagaa 360  
 Db 430 ggaatcctttttagcttgcagagtggttgcaaatcattcattcactccttattcttgagaa 489  
 OY 361 actcctgttcttgcagtgctcctcagtgaggaagagtgatagtgaggaagga 420  
 Db 490 actcctgttcttgcagtgctcctcagtgaggaagagtgatagtgaggaagga 549  
 OY 421 aactcagtggttgaagaccttgcagtgctcctcagtgaggaagagtgatagtgagga 480  
 Db 550 aactcagtggttgaagaccttgcagtgctcctcagtgaggaagagtgatagtgagga 609  
 OY 481 gaaaactcttctcagtgctcctcagtgaggaagagtgatagtgaggaagagtgac 540  
 Db 610 gaaaactcttctcagtgctcctcagtgaggaagagtgatagtgaggaagagtgac 669  
 OY 541 ctgccttcttcttgcagtgctcctcagtgaggaagagtgatagtgaggaagagtgac 600  
 Db 670 ctgccttcttcttgcagtgctcctcagtgaggaagagtgatagtgaggaagagtgac 729  
 OY 601 aagcattcagcagaagatcattcctcgtattatatactcagtgagcctggcaggtttgat 660  
 Db 730 aagcattcagcagaagatcattcctcgtattatatactcagtgagcctggcaggtttgat 789  
 OY 661 gaatttggaagcgttatgttggaagaccttggaacattcagaagatccttctaagacct 720  
 Db 790 gaatttggaagcgttatgttggaagaccttggaacattcagaagatccttctaagacct 849  
 OY 721 gttagcctctgcgaagaatttgcagatgacatgtacagttcttattatgttggaagtgacgt 780  
 Db 850 gttagcctctgcgaagaatttgcagatgacatgtacagttcttattatgttggaagtgacgt 909  
 OY 781 gttagagtgtagtcaagtgatcatttggaacacctcctcatttggaagcagaagacatc 840  
 Db 910 gttagagtgtagtcaagtgatcatttggaacacctcctcatttggaagcagaagacatc 969  
 OY 841 cttagaggaagaagaagcagaagacccaagaatcctcattatcacttgcatalatagataat 900  
 Db 970 cttagaggaagaagaagcagaagacccaagaatcctcattatcacttgcatalatagataat 1029  
 OY 901 tttagaatattcgttggttttcaacatgtaacttggataatgcttgcttgcttgct 960  
 Db 1030 tttagaatattcgttggttttcaacatgtaacttggataatgcttgcttgcttgct 1089  
 OY 961 gtgattatcactcttaacatatttggaacatgagcttgatagatagatcattat 1020  
 Db 1090 gtgattatcactcttaacatatttggaacatgagcttgatagatagatcattat 1149  
 OY 1021 aggatgacaacacgaagattcgaatgat 1050







```
Query Match      28.2%; Score 296; DB 22; Length 560;
Best Local Similarity 100.0%; Pred. No. 1.8e-140;
Matches 296; Conservative 0; Mismatches 0; Indels 0; Gaps 0
```

QY 1 atggcgtgtgtttgacgtgtccctccggaggtttgtatgacgagtgatttgc 60  
Db 162 atggcgtgtgtttgacgtgtccctccggaggtttgtatgacgagtgatttgc 221  
QY 61 atattaaatccaccggagctgtctttttccgaaatggaattgtccctacacgaag 120  
Db 221 atattaaatccaccggagctgtctttttccgaaatggaattgtccctacacgaag 281

Db	QY	QY	Db
282	cgagtcaccagacg	cgagtcaccagacg	cgagtcaccagacg
181	ccagagactcgacg	ccagagactcgacg	ccagagactcgacg
342	ccagagactcgacg	ccagagactcgacg	ccagagactcgacg
241	aagagagtgtaaca	aagagagtgtaaca	aagagagtgtaaca
402	aagagagtgtaaca	aagagagtgtaaca	aagagagtgtaaca

RESULT	8
AAx51918	
ID	AAx51918 standard; DNA; 343 BP
XX	

DT 22-JUN-1999 (first entry)  
XX

Human secreted protein 5' EST SEQ ID NO: 132.

human; secreted protein; EST; expressed sequence tag; diagnosis; forensic; gene therapy; chromosome mapping; signal peptide; upstream regulatory sequence; cytokine activity; cell proliferation; differentiation; haematopoiesis regulation; tissue growth regulation; reproductive hormone regulation; chemotactic; chemokinetic; haemostatic; thrombolytic; anti-inflammatory; tumour inhibition; ds.

00S	Homo sapiens.	
xxx		
00PN	W09060552-A2.	
xxx		
00PD	11-FEB-1999.	
xxx		
00PF	31-JUL-1998;	98WO-1B01236
xxx		
00PR	01-AUG-1997;	97US-0905223
xxx		

(GEST ) GENSET

Duclet A, Dumas Milne Edwards J, Lacroix B,

WPI; 1999-153782/13.  
P-PSDB: AAV12110

P-PSDB; AAY13118.

new isolated brain-derived nucleic acids - used to develop products which may have cytokine, immune, regulatory, haematopoiesis regulating, anti-inflammatory or tumour inhibition activity

XX Claim 1; Page 273-274; 577pp; English  
PS  
XX

AAV31/8 to AAV32019 represent 5' expressed sequence tags (ESTs) for human secreted proteins, and encode the proteins given in AAV12967 to AAV13219, respectively. The proteins given represent the signal peptide and an N-terminal fragment of a secreted protein. The nucleic acid sequences can be used for producing secreted human gene products. They can also be used to develop products for diagnosis and therapy. The proteins obtained may have cytokine activity, cell proliferation/differentiation activity, and so on.

CC Sequence 343 BP; 56 A; 88 C; 104 G; 88 T; 7 other;  
CC activity, tissue growth regulating activity, haemadapoptosis regulating  
CC activity, tissue growth regulating activity, reproductive hormone  
CC regulating activity, chemotactic, chemokinetic activity, haemostatic and  
CC thrombolytic activity, receptor / ligand activity, anti-inflammatory  
CC activity, tumour inhibition activity or other activities. The products  
CC can be used in forensic, gene therapy and chromosome mapping procedures.  
CC The sequences can also be used for obtaining corresponding promoter  
CC sequences. The nucleic acids encoding the signal peptide can be used for  
CC directing extracellular secretion of a polypeptide or the insertion of a  
CC polypeptide into a membrane, or importing a polypeptide into a cell.  
XX

Query Match	21.28;	Score 223;	DB 20;	Length 343;
Best Local Similarity	100.0%;	Pred. No. 2.1e-103;		
Matches 223;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

[illegible]

RESULT	9
AAZ12587	
ID	AAZ12587 standard; cDNA; 254 BP.
XY	

AAZ12587;

DT	12-OCT-1999	(first entry)
XX		

Human gene	expression	product	CDNA sequence	SEQ ID NO:56
XX				

human; gene; gene expression product; diagnosis; therapy; probe;  
 detection; mapping; tissue typing; profiling; forensic; cancer;  
 genetic analysis; colorectal cancer; breast cancer; lung cancer; ss

Homo sapiens

W099938972-A2

05-AUG-1999

28-JAN-1999; 99WO-US01619.

03-APR-1998; 98US-0080666.  
28-JAN-1999; 08HC 0073010

24-FEB-1998; 98US-0075954.  
31-MAR-1998; 98US-0080114

03-APR-1998; 98US-0080515.

XX

PA (CHIR ) CHIRON CORP.  
 PA (HYSE-) HYSEQ INC.  
 XX  
 XX  
 PI Cirvenjakov R, Dickson M, Drmanac R, Drmanac S;  
 PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis M;  
 PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I,  
 PI Lamsom G, Leshkowitz D, Pot D, Randazzo F, Reinhard C,  
 PI Stache-Crain B, Sudduth-Klinger J, Williams LT;  
 XX  
 DR WPI: 1999-494092/41.  
 PT Novel human genes and their expression products which are  
 PT differentially expressed in different cell types  
 XX  
 PS Claim 1; Page 674; 2479pp; English.  
 XX  
 CC The present invention describes a library of human polynucleotides  
 CC comprising the sequences given in AA212532 to AA21779. Also described is  
 CC a method of detecting differentially expressed genes correlated with the  
 CC cancerous state of a mammalian cell, comprising detecting at least one  
 CC differentially expressed gene product in a test sample from a cell  
 CC suspected of being cancerous, where the gene product is encoded by one  
 CC of the 5248 polynucleotide sequences given in AA212532 to AA21779. The  
 CC polynucleotides can be used as a source of primers and probes, which can  
 CC be used for a variety of purpose, e.g. detection of expression levels,  
 CC mapping, tissue typing or profiling, forensics, genetic analysis and  
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides  
 CC can be used for raising antibodies for experimental, diagnostic and  
 CC therapeutic purposes. The polynucleotides may also be used to construct  
 CC arrays for diagnostics (which may be used to determine function of an  
 CC encoded protein); and to detect differences in expression levels between  
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to  
 CC identify a genetic predisposition or susceptibility to a disease such as  
 CC cancer). The polynucleotides of the invention are especially used in the  
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,  
 CC and lung cancer. The polynucleotides can also be used to screen for  
 CC peptide analogues and antagonists.  
 XX  
 SQ Sequence 254 BP; 39 A; 50 C; 55 G; 63 T; 47 other;  
 XX  
 Query Match 14.4%; Score 151; DB 20; Length 254;  
 Best Local Similarity 100.0%; Pred. No. 7.7e-67;  
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 97 ggaattgctataccagagagagatcccgagatgctgctgcatgtccatggcttc 156  
 Db 1 ggaattgctataccagagagagatcccgagatgctgctgcatgtccatggcttc 60  
 QY 157 tctgtgaagaagacattcttggccagagactcgagtggtgtaacctgttcact 216  
 Db 61 tctgtgaagaagacattcttggccagagactcgagtggtgtaacctgttcact 120  
 QY 217 cgggctaccgtcatgtgtagtgaaggag 247  
 Db 121 cgggctaccgtcatgtgtagtgaaggag 151  
 RESULT 10  
 AA215793  
 ID AA215793 standard; cDNA; 858 BP.  
 XX  
 AC AA215793;  
 XX  
 DT 12-OCT-1999 (first entry)  
 XX  
 DE Human gene expression product cDNA sequence SEQ ID NO:3262.  
 XX  
 XX Human; gene: gene expression product; diagnosis; therapy; probe;  
 KW detection; mapping; tissue typing; profiling; forensics; cancer;  
 XX genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.  
 OS Homo sapiens.

---

XX  
 PN W09938972-A2.  
 XX  
 XX 05-AUG-1999.  
 XX  
 XX 28-JAN-1999; 99WO-US01619.  
 XX  
 PF 03-APR-1998; 98US-0080666.  
 XX 28-JAN-1998; 98US-0072910.  
 PR 24-FEB-1998; 98US-0075954.  
 PR 31-MAR-1998; 98US-0080114.  
 PR 03-APR-1998; 98US-0080515.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (HYSE-) HYSEQ INC.  
 XX  
 XX  
 PI Cirvenjakov R, Dickson M, Drmanac R, Drmanac S;  
 PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis M;  
 PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I,  
 PI Lamsom G, Leshkowitz D, Pot D, Randazzo F, Reinhard C,  
 PI Stache-Crain B, Sudduth-Klinger J, Williams LT;  
 XX  
 DR WPI: 1999-494092/41.  
 PT Novel human genes and their expression products which are  
 PT differentially expressed in different cell types  
 XX  
 PS Claim 1; Page 1565; 2479pp; English.  
 XX  
 CC The present invention describes a library of human polynucleotides  
 CC comprising the sequences given in AA212532 to AA21779. Also described is  
 CC a method of detecting differentially expressed genes correlated with the  
 CC cancerous state of a mammalian cell, comprising detecting at least one  
 CC differentially expressed gene product in a test sample from a cell  
 CC suspected of being cancerous, where the gene product is encoded by one  
 CC of the 5248 polynucleotide sequences given in AA212532 to AA21779. The  
 CC polynucleotides can be used as a source of primers and probes, which can  
 CC be used for a variety of purpose, e.g. detection of expression levels,  
 CC mapping, tissue typing or profiling, forensics, genetic analysis and  
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides  
 CC can be used for raising antibodies for experimental, diagnostic and  
 CC therapeutic purposes. The polynucleotides may also be used to construct  
 CC arrays for diagnostics (which may be used to determine function of an  
 CC encoded protein); and to detect differences in expression levels between  
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to  
 CC identify a genetic predisposition or susceptibility to a disease such as  
 CC cancer). The polynucleotides of the invention are especially used in the  
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,  
 CC and lung cancer. The polynucleotides can also be used to screen for  
 CC peptide analogues and antagonists.  
 XX  
 SQ Sequence 858 BP; 205 A; 186 C; 189 G; 225 T; 53 other;  
 XX  
 Query Match 13.0%; Score 137; DB 20; Length 858;  
 Best Local Similarity 100.0%; Pred. No. 9.7e-60;  
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 316 ctgtgaagtgttgaataattctacccttattcttgaggaacccgtgtgtttg 375  
 Db 283 ctgtgaagtgttgaataattctacccttattcttgaggaacccgtgtgtttg 342  
 QY 376 caatttgctcccaattgaggaagtgtatattgtaaggagcaactcagtgattga 435  
 Db 343 cagttgctcccaattgaggaagtgtatattgtaaggagcaactcagtgattga 402  
 QY 436 gaccttgagttacctt 452  
 Db 403 gaccttgagttacctt 419  
 RESULT 11  
 AA57151

ID AAA57151 standard; cDNA; 877 BP.  
 AC AAA57151;  
 XX  
 DT 03-OCT-2000 (first entry)  
 XX  
 DE Mouse Pox CDNA.  
 XX  
 KW Mouse; Pox; homeobox; placenta-specific; placental development;  
 KW gestational trophoblastic disease; cytosolic; transcriptional regulator;  
 KW abortion induction; gestational trophoblastic tumor;  
 KW invasive mole; chorioncarcinoma; placental site trophoblastic tumor; ss.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 44..787  
 FT /\*tag= a  
 FT /product= "Pox protein"  
 XX  
 PN US6063912-A.  
 XX  
 PD 16-MAY-2000.  
 XX  
 PE 06-AUG-1998; 98US-0129888.  
 XX  
 PR 06-AUG-1998; 98US-0129888.  
 XX  
 PA (KOKU-) KOREA KUMHO PETROCHEMICAL CO LTD.  
 XX  
 PI Chun J, Han Y;  
 XX  
 DR WPI: 2000-375541/32.  
 DR P-PSDB; AAY99886.  
 XX  
 PT New nucleic acid encoding placenta homeobox protein (Pox) useful for  
 PT developing treatment for trophoblast-specific Pox-mediated diseases,  
 PT particularly gestational trophoblastic diseases, e.g. gestational  
 PT trophoblastic tumors -  
 XX  
 PS Claim 5; Fig 1; 23pp; English.  
 XX  
 CC The present sequence encodes the mouse Pox protein, a novel  
 CC homeodomain-containing transcription factor. The sequence was isolated  
 CC by screening a 13.5 day-old mouse conceptus cDNA library for  
 CC homeobox-containing sequences using a PCR cloning strategy. Expression  
 CC of the Pox polynucleotide is restricted to the placenta, particularly  
 CC to the placenta trophoblast cell layers during embryogenesis. The  
 CC expression pattern of Pox may be exploited to detect trophoblast specific  
 CC lineages, such as the labyrinthine trophoblast layer and giant cells. The  
 CC Pox gene and gene product may also be used to develop treatment for  
 CC trophoblast-specific Pox-mediated diseases, such as gestational  
 CC trophoblastic diseases (e.g. gestational trophoblastic tumours, invasive  
 CC moles, chorioncarcinomas, and placental site trophoblastic tumours).  
 CC They may also be used to identify putative abortion-inducing agents,  
 CC which may offer an alternative to surgical abortion.  
 CC  
 XX  
 SQ Sequence 877 BP; 251 A; 204 C; 256 G; 166 T; 0 other;

Query Match 1.7%; Score 18; DB 21; Length 877;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 765 tgggtggaatgcagtggt 782  
 |||  
 DB 124 tgggtggaatgcagtggt 141

RESULT 12  
 AAC99977/c  
 ID AAC99977 standard; cDNA; 1274 BP.  
 XX

AC AAC99977;  
 XX  
 DT 13-MAR-2001 (first entry)  
 XX  
 DE Human secreted protein gene 142 SEQ ID NO:170.  
 XX  
 KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;  
 KW antirheumatic; antiproliferative; cytostatic; cardiac; vasotropic;  
 KW cerebroprotective; nootropic; neuroprotective; antibacterial; vitruicide;  
 KW fungicide; ophthalmological; gene therapy; pathological condition;  
 KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;  
 KW neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;  
 KW cerebrovascular disorder; angiogenesis; nervous system disorder;  
 KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
 KW wound healing; skin aging; food additive; preservative;  
 KW chromosome 20; ss.  
 OS  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200070042-A1.  
 XX  
 PD 23-NOV-2000.  
 XX  
 PE 11-MAY-2000; 2000MO-US12788.  
 XX  
 PR 13-MAY-1999; 99US-0134068.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CE;  
 PI Duan RD, Florence KA, Soppet DR;  
 XX  
 DR WPI: 2000-679828/66.  
 DR P-PSDB; AAB56236.  
 XX  
 PT Isolated nucleic acid molecule encoding a human secreted protein is  
 PT used in preventing, treating or ameliorating a medical condition -  
 XX  
 PS Claim 1; Page 939-940; 1065pp; English.  
 XX  
 CC The polynucleotide sequences given in AAC99818 to AAC99977 encode the  
 CC human secreted proteins given in AAB56077 to AAB56362. Human secreted  
 CC proteins have activities based on the tissues and cells the genes are  
 CC expressed in. Examples of activities include: immunosuppressive;  
 CC antirheumatic; antineumatic; antiproliferative; cytostatic; cardiac;  
 CC vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;  
 CC vitruicide; fungicide; and ophthalmological. The human secreted  
 CC polynucleotides and proteins can be used to prevent, treat or ameliorate  
 CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,  
 CC dogs, chickens or sheep. They are also used in diagnosing a pathological  
 CC condition or susceptibility to a pathological condition. Disorders which  
 CC are diagnosed or treated include autoimmune diseases e.g. rheumatoid  
 CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or  
 CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular  
 CC disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders  
 CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and  
 CC fungi and ocular disorders e.g. corneal infection. The proteins can also  
 CC be used to aid wound healing and epithelial cell proliferation, to  
 CC prevent skin aging due to sunburn, to maintain organs before  
 CC transplantation, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The proteins can also be used as a  
 CC food additive or preservative to increase or decrease storage  
 CC capabilities. AAC99809 to AAC99817 and AAB56076 represent sequences used  
 CC in the exemplification of the present invention.  
 CC  
 XX  
 SQ Sequence 1274 BP; 390 A; 256 C; 271 G; 357 T; 0 other;

Query Match 1.7%; Score 18; DB 21; Length 1274;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 267 ccagagcagtgcatc 284







CC proteins stimulating expression of cell cycle dependent kinase  
 CC inhibitors and arresting cell cycle progression at specific boundaries  
 CC to thereby modulate cell proliferation. As p62 acts to boost B cell  
 CC response and may be used to treat disorders where this is beneficial,  
 CC e.g. infections by pathogenic microorganisms, e.g. bacteria, viruses and  
 CC protozoans. p62 can be used to expand T cell populations for treating  
 CC infectious diseases or cancer, e.g. the resulting cells may be transduced  
 CC to render them resistant to HIV infection. Inhibitors of p62 can be used  
 CC to reduce B or T cell responses and may be used to treat a variety of  
 CC autoimmune diseases, e.g. diabetes mellitus, arthritis, multiple  
 CC sclerosis allergic reactions, Crohn's diseases etc.

SO Sequence 2083 BP; 447 A; 588 C; 608 G; 440 T; 0 other;

Query Match 1.7%; Score 18; DB 18; Length 2083;

Best Local Similarity 100.0%; Pred. No. 26;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 143 tgtccatggctctctcgt 160

DB 1271 tgtccatggctctctcgt 1288

RESULT 16

AA16632/c

ID AA16632 standard; cDNA; 2472 BP.

AC AA16632;

DT 16-JUN-2000 (first entry)

DE Human secreted protein clone pp314\_19 nucleotide sequence SEQ ID NO:29.

XX Human: secreted protein; immunostimulant; immunosuppressant; vitruicide;  
 KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;  
 KW antidiabetic; antihistaminic; antirheumatic; antirheumatic; protozoacide;  
 KW antihydroid; immune deficiency; severe combined immunodeficiency; SCID;  
 KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;  
 KW connective tissue disease; multiple sclerosis; erythematosis;  
 KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;  
 KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;  
 KW insulin dependent diabetes mellitus; graft-versus-host-disease;  
 KW autoimmune inflammatory eye disease; allergy; ss.

OS Homo sapiens.

XX W0200009552-A1.

PD 24-FEB-2000.

PF 13-AUG-1999; 99W0-US18298.

PR 14-AUG-1998; 98US-0096622.

PR 17-AUG-1998; 98US-0096615.

PR 04-SEP-1998; 98US-0099329.

PR 23-OCT-1998; 98US-0105368.

PR 08-JAN-1999; 99US-0115234.

PR 12-FEB-1999; 99US-0119931.

PR 18-FEB-1999; 99US-0120575.

PR 30-APR-1999; 99US-0132020.

PR 11-AUG-1999; 99US-0096622.

PT New polynucleotides encoding secreted proteins, which may have e.g.

PT nutritional, chemokine, immune stimulating or suppressing,  
 PT hematopoiesis regulating, tissue growth, activity/inhibin  
 PT antiinflammatory or tumor inhibition activity

PS Claim 38; Page 498-499; 641pp; English.

XX AA16618 to AA16697 encode the human secreted proteins given in  
 CC AA16618 to AA16697 isolated from human adult brain, adult thyroid,  
 CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,  
 CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,  
 CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,  
 CC and adult bladder, cDNA libraries. The polynucleotides and proteins are  
 CC predicted to have biological activities which would make them suitable  
 CC for treating, preventing or ameliorating medical conditions in humans  
 CC and animals. The polynucleotides can be used as markers for tissues in  
 CC which the protein is preferentially expressed, as molecular weight  
 CC markers on Southern gels, and as chromosome markers or tags to identify  
 CC chromosomes or to map gene positions. The proteins can be used in the  
 CC treatment of immune deficiencies and disorders, such as severe combined  
 CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other  
 CC infections. These infections include human immunodeficiency virus (HIV),  
 CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and  
 CC candidiasis. The proteins can be used to treat autoimmune disorders such  
 CC as connective tissue disease, multiple sclerosis, systemic lupus  
 CC erythematosis, rheumatoid arthritis, autoimmune pulmonary inflammation,  
 CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent  
 CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and  
 CC autoimmune inflammatory eye disease. The proteins can also be used to  
 CC treat allergic conditions, such as asthma. AA16698 to AA16774 represent  
 CC probes for the human secreted proteins from the present invention.

SO Sequence 2472 BP; 796 A; 475 C; 502 G; 699 T; 0 other;

Query Match 1.7%; Score 18; DB 21; Length 2472;

Best Local Similarity 100.0%; Pred. No. 26;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 267 ccagcagcagtccttc 284

DB 1778 CCAGCAGCAGTCCTTC 1761

RESULT 17

AAV62145/c

ID AAV62145 standard; DNA; 2694 BP.

AC AAV62145;

DT 07-DEC-1998 (first entry)

DE HSV-2 strain SB5 Contig ID 90 DNA sequence.

XX HSV-2 strain SB5; immunological response induction; therapy;

KW antiviral identification; viral protein inhibitor; ss.

OS Herpes simplex virus type 2.

XX Key Location/Qualifiers

FT CDS 501..2694

FT /tag= a

FT /transl\_except= (pos:2694, aa:Arg)

FT /note= "no stop codon given"

PT W09820016-A1.

PD 14-MAY-1998.

PF 31-OCT-1997; 97W0-US20016.

PR 09-JUN-1997; 97US-0049018.

PR 04-NOV-1996; 96US-0030279.

PA (SMIK ) SMITHKLINE BEECHAM CORP.

PI Chan JY, Dabrowski-Amarel CE, Delvecchio AM, Dillon SB;

PI Esser KM, Leary JF;

DR WPI: 1998-286847/25.

DR P-PSDB: AAW72061.

PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention  
PT and treatment of infection or inducing immunological response in  
PT mammal

PS Claim 1: Page 280-281; 748pp; English.

CC This sequence represents a Herpes simplex virus type-2 (HSV-2) DNA  
CC sequence of the invention. This sequence was isolated from HSV-2 strain  
CC 585 (deposited as ATCC VR-2546), is designated Conlig ID 90, and encodes  
CC a HSV-2 protein. The proteins can be used for the treatment or  
CC prevention of disease, to induce an immunological response in a mammal or  
CC to identify inhibitors, activators or novel antivirals. Antagonists of  
CC the proteins can be used to inhibit a viral polypeptide. The DNA sequence  
CC or a vector containing it can also be used to induce an immunological  
CC response in a mammal.

CC Sequence 2694 BP; 464 A; 953 G; 910 G; 365 T; 2 other;

Query Match 1.7%; Score 18; DB 19; Length 2694;

Best Local Similarity 100.0%; Pred. No. 26;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 333 ttccattcactcccttatt 350

DB 70 ttccattcactcccttatt 53

RESULT 18

AAA08567/c

ID AAA08567 standard; DNA; 2776 BP.

XX AAA08567;

DT 19-JUL-2000 (first entry)

DE Human chaperone protein 5 (HCHP-5) coding sequence.

KW Human chaperone protein; HCHP-5; neurodegenerative; metabolic;

KW developmental; autoimmune; inflammatory; cell proliferation; cancer; ss.

OS Homo sapiens.

PH Key Location/Qualifiers

FT CDS 748..2460

FT /tag= a

FT /product= HCHP-5

PN MO200017358-A2.

PD 30-MAR-2000.

PF 22-SEP-1999; 99WO-US22027.

PR 22-SEP-1998; 98US-0172221.

PR 19-JAN-1999; 99US-0233291.

PR 19-APR-1999; 99US-0172232.

PA (INCY-) INCYTE PHARM INC.

PI Tang YT, Hillman JL, Yue H, Patterson C, Baughn MR, Batra S;

DR WPI: 2000-283583/24.

DR P-PSDB: AAY91944.

PT New purified polypeptides and polynucleotides encoding human chaperone  
PT proteins, useful for diagnosing, treating and preventing disorders  
PT associated with the expression human chaperone proteins

PS Claim 7: Page 79-80; 88pp; English.

CC AAA08563-68 encode human chaperone proteins 1-6 (HCHP-1 to HCHP-6)  
CC respectively. The sequences can be used to treat and prevent disorders  
CC associated with altered expression or activity of HCHP comprising  
CC administering a composition comprising the polypeptide or an antagonist  
CC to a patient (claimed). The human chaperone proteins are also useful for  
CC the diagnosis, treatment or prevention of neurodegenerative, metabolic,  
CC developmental, autoimmune/inflammatory disorders and cell proliferative  
CC disorders including cancer.

CC Sequence 2776 BP; 834 A; 535 G; 814 T; 0 other;

Query Match 1.7%; Score 18; DB 21; Length 2776;

Best Local Similarity 100.0%; Pred. No. 26;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 267 cccagcagctgctatc 284

DB 1897 CCCAGCAGCTGCTATTC 1880

RESULT 19

AAA08113

ID AAA08113 standard; DNA; 4085 BP.

XX AAA08113;

DT 22-JUN-2000 (first entry)

DE Human bladder specific gene 232110 clone 378221H1 SEQ ID NO:3.

KW Human; bladder specific gene; BSG; bladder cancer; diagnosis;

KW detection; metastasis; cytostatic; ds.

OS Homo sapiens.

PN MO200012761-A1.

PD 09-MAR-2000.

PF 01-SEP-1999; 99WO-US20003.

PR 02-SEP-1998; 98US-0098825.

PA (DIAD-) DIADEXUS LLC.

PI Salceda S, Sun Y, Recipon H, Caferkey R;

DR WPI: 2000-256658/22.

PT Detecting, diagnosing metastasis, staging, monitoring, imaging and  
PT treating bladder cancer, involves measuring bladder specific gene (BSG)  
PT levels in cells or body fluids

PS Example 1: Page 27-28; 31pp; English.

CC A method has been developed for detecting, diagnosing metastasis,  
CC staging and monitoring the onset of metastasis or a change in stage  
CC of bladder cancer, by measuring bladder specific gene (BSG) levels in  
CC cells, tissues or body fluids. An increase in BSG levels compared to  
CC normal human control is associated with the presence or progression of  
CC bladder cancer; a decrease is associated with cancer regression. The  
CC method comprising measuring BSG is useful for detecting, diagnosing  
CC metastasis, staging and monitoring the onset of metastasis or a change  
CC in stage of bladder cancer. An antibody against BSG (I) labeled with  
CC paramagnetic ions or an isotope is useful for imaging bladder cancer  
CC and (II), especially conjugated to a cytotoxic agent is useful for

treating bladder cancer. The method is efficient in discriminating between bladder cancer which has metastasised and the bladder cancer which has not metastasised. The present sequence represents a BSG sequence used in the exemplification of the present invention.

Sequence 4085 BP; 943 A; 1089 C; 1173 G; 879 T; 1 other;

Query Match 1.7%; Score 18; DB 21; Length 4085;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
143 tttccatgggtctctctg 160  
|||||  
Db 1238 ttccatgggtctctctg 1255

RESULT 20  
AAZ47589/C  
ID AAZ47589 standard; DNA; 80 BP.

AC AAZ47589;

DT 14-MAR-2000 (first entry)

DE PCR primer used in the production of endostatin-NC1 trimer.

XX Endostatin; scatter factor activity; human; tubulogenesis; psoriasis;

KW metastatic cancer; tumorigenesis; ocular angiogenic disease;

KM rheumatoid arthritis; Osler-Webber syndrome; telangiectasia;

KW haemophilic joint; angiodioma; wound granulation; PCR primer; ss.

XX Synthetic.

OS Homo sapiens.

XX WO962944-A2.

PD 09-DEC-1999.

XX 03-JUN-1999; 99MO-US12278.

XX 03-JUN-1998; 98US-0087890.

PR 10-JUL-1998; 98US-0092393.

PR 01-SEP-1998; 98US-0098790.

XX (CHIL-) CHILDRENS MEDICAL CENT.

PI Javaherian K, Folkman MJ;

DR WPI: 2000-072833/06.

XX New endostatin oligomers, used for treating e.g. tumours -

PS Example 1; Page 23; 44pp; English.

XX PCR primers AAZ47589-247590 are used to amplify a section of human  
CC collagen 18, in the production of endostatin-NC1 trimers. Endostatin is  
CC an approximately 20kD C-terminal globular domain of the collagen-like  
CC protein collagen XVIII. Protein oligomers consisting of more than one  
CC endostatin monomer have anti-tubulogenic effects and induce  
CC reorganization of the actin cytoskeleton. The oligomer has scatter  
CC factor activity. The oligomers induce the destruction of tubular lumens  
CC and elongation of cells, and inhibit tubulogenesis and tumorigenesis.  
CC The oligomers can also be used to treat metastatic cancers, tumours,  
CC rheumatoid arthritis, psoriasis, ocular angiogenic disease, Osler-Webber  
CC syndrome, plaque neovascularisation, telangiectasia, haemophilic  
CC joints, angiodioma and wound granulation. The oligomers can also be  
CC used to treat diseases that have angiogenesis as a pathological  
CC consequence e.g. ulcers. The endostatin oligomers can also be used to  
CC develop affinity columns for isolating antibodies or receptors. Passive  
CC antibody therapy using antibodies that specifically bind endostatin  
CC oligomers can be used to modulate morphogenic processes such as  
CC metastatic cancer as well as angiogenesis-dependent processes such as

reproduction, development, wound healing, tissue repair, and angiogenesis-dependent diseases. Also, antisera directed to the Fab regions of endostatin oligomer antibodies can be administered to block the ability of endogenous endostatin oligomer antisera to bind endostatin oligomers.

Sequence 80 BP; 20 A; 26 C; 23 G; 11 T; 0 other;

Query Match 1.6%; Score 17; DB 21; Length 80;  
Best Local Similarity 100.0%; Pred. No. 88;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
225 cgtcatgctgctgctga 241  
|||||  
Db 39 cgtcatgctgctgctga 23

RESULT 21  
AAA61693/C  
ID AAA61693 standard; DNA; 117 BP.

AC AAA61693;

DT 23-OCT-2000 (first entry)

DE pTrypHis plasmid oligonucleotide, SEQ ID NO:40.

XX BSSP2; serine protease; mouse; mBSSP2; human; hbSSP2; brain;

KW diagnostic marker; antibody; transgenic animal; Alzheimer's disease;

KW epilepsy; cancer; inflammation; inflammatory disorder; infertility;

KW prostatic hypertrophy; plasmid construction; ss.

XX Synthetic.

OS WO200031272-A1.

PD 02-JUN-2000.

XX 19-NOV-1999; 99MO-JP06475.

XX 20-NOV-1998; 98JP-0347785.

PA (FUSO) FUSO PHARM IND LTD.

PI Uemura H, Okui A, Komimami K, Yamaguchi N, Mitsui S;

DR WPI: 2000-400082/34.

XX Serine protease BSSP2, useful in detecting homologs, mutants and  
PT polymorphic variants as markers for diagnosis of e.g. Alzheimer's  
PT disease, cancer, inflammation and prostate hypertrophy, using blood,  
PT urine or other tissues -

PS Example 4; Page 89; 92pp; Japanese.

XX The invention relates to novel serine proteases designated BSSP2  
CC (AA61655-B1669), and to nucleic acids encoding them (AAA61659-A61663).  
CC The invention also relates to vectors and transformants comprising BSSP2  
CC nucleic acids; transgenic animals in which the expression level of BSSP2  
CC can be varied; and an mBSSP2 knockout mouse. The invention additionally  
CC encompasses anti-BSSP2 antibodies and methods of production of such  
CC antibodies, methods of BSSP2 detection using the antibodies, and the  
CC use of BSSP2 proteins or fragments as diagnostic markers for certain  
CC medical conditions. Nucleotides encoding BSSP2 were initially isolated  
CC in a mouse brain cDNA library using degenerate PCR primers (AAA61673-  
CC AAA61674) based on conserved regions of serine proteases. The BSSP2  
CC serine proteases and nucleotides encoding them are useful in detecting  
CC homologs, mutants and polymorphic variants in biological samples  
CC (e.g., blood, urine, brain, prostate gland and testis) as diagnostic  
CC markers for conditions such as Alzheimer's disease, epilepsy, cancer,  
CC inflammation, infertility and prostatic hypertrophy. Sequences  
CC AAA61664-A61665 and AAA61693-A61694 represent oligonucleotides used to

CC construct plasmids used in the invention.  
XX  
SQ Sequence 117 BP; 27 A; 38 C; 20 G; 32 T; 0 other;

Query Match 1.6%; Score 17; DB 21; Length 117;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 225 cgcacatgctgacgtgca 241  
|||||  
Db 95 CGTCATGCTGATGCTGA 79

## RESULT 22

AAA61694  
ID AAA61694 standard; DNA; 117 BP.

AC AAA61694;

DT 23-OCT-2000 (first entry)

DE pTrypHis plasmid oligonucleotide, SEQ ID NO:41.

XX BSSP2; serine protease; mouse; mBSSP2; human; hBSSP2; brain;  
KW diagnostic marker; antibody; transgenic animal; Alzheimer's disease;  
KW epilepsy; cancer; inflammation; inflammatory disorder; infertility;  
KW prostatic hypertrophy; plasmid construction; ss.

OS Synthetic.

PN WO200031272-A1.

PD 02-JUN-2000.

PF 19-NOV-1999; 99WO-JP06475.

PR 20-NOV-1998; 98JP-0347785.

PA (FUSO ) FUSO PHARM IND LTD.

PI Uemura H, Okui A, Komitani K, Yamaguchi N, Mitsui S;

WPI: 2000-400082/34.

PT Serine protease BSSP2, useful in detecting homologs, mutants and  
PT polymorphic variants as markers for diagnosis of e.g. Alzheimer's  
PT disease, cancer, inflammation and prostate hypertrophy, using blood,  
PT urine or other tissues -

PS Example 4; Page 89; 92pp; Japanese.

XX The invention relates to novel serine proteases designated BSSP2  
XX (AA61693-B11699), and to nucleic acids encoding them (AAA61695-A61663)  
CC The invention also relates to vectors and transformations comprising BSSP2  
CC nucleic acids; transgenic animals in which the expression level of BSSP2  
CC can be varied; and an mBSSP2 knockout mouse. The invention additionally  
CC encompasses anti-BSSP2 antibodies and methods of production of such  
CC antibodies, methods of BSSP2 detection using the antibodies, and the  
CC use of BSSP2 proteins or fragments as diagnostic markers for certain  
CC medical conditions. Nucleotides encoding BSSP2 were initially isolated  
CC in a mouse brain cDNA library using degenerate PCR primers (AAA61673-  
CC AAA61674) based on conserved regions of serine proteases. The BSSP2  
CC serine proteases and nucleotides encoding them are useful in detecting  
CC homologues, mutants and polymorphic variants in biological samples  
CC (e.g., blood, urine, brain, prostate gland and testis) as diagnostic  
CC markers for conditions such as Alzheimer's disease, epilepsy, cancer,  
CC inflammation, infertility and prostatic hypertrophy. Sequences  
CC AAA61664-A61665 and AAA61693-A61694 represent oligonucleotides used to  
CC construct plasmids used in the invention.

SQ Sequence 117 BP; 32 A; 20 C; 38 G; 27 T; 0 other;

Query Match 1.6%; Score 17; DB 21; Length 117;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 225 cgcacatgctgacgtgca 241  
|||||  
Db 23 cgcacatgctgacgtgca 39

## RESULT 23

AAA61731/C  
ID AAA61731 standard; DNA; 117 BP.

AC AAA61731;

DT 23-OCT-2000 (first entry)

DE pTrypHis plasmid oligonucleotide, SEQ ID NO:47.

XX BSSP4; serine protease; human; hBSSP4; mouse; mBSSP4; brain;  
KW diagnostic marker; antibody; transgenic animal; Alzheimer's disease;  
KW oedema; dropsy; cancer; inflammation; prostate; testis; bone;  
KW plasmid construction; ss.

OS Synthetic.

PN WO200031277-A1.

PD 02-JUN-2000.

PF 19-NOV-1999; 99WO-JP06472.

PR 20-NOV-1998; 98JP-0347813.

PA (FUSO ) FUSO PHARM IND LTD.

PI Uemura H, Okui A, Komitani K, Yamaguchi N, Mitsui S;

WPI: 2000-400084/34.

PT Serine protease BSSP4 and antibodies recognizing BSSP4 for assay and  
PT diagnosis of diseases in which BSSP4 expression is altered -

PS Example 4; Page 105; 111pp; Japanese.

XX The invention relates to novel serine proteases designated BSSP4  
XX (AA61700-B11709), and to nucleic acids encoding them (AAA61695-A61704,  
CC AAA61799). The invention also relates to vectors and transformations  
CC comprising BSSP4 nucleic acids; transgenic animals in which the  
CC expression level of BSSP4 can be varied; and an mBSSP4 knockout mouse.  
CC The invention additionally encompasses anti-BSSP4 antibodies and methods  
CC of production of such antibodies, methods of BSSP4 detection using the  
CC antibodies, and the use of BSSP4 proteins or fragments as diagnostic  
CC markers for certain medical conditions. Nucleotides encoding BSSP4 were  
CC initially isolated in a human brain cDNA library using degenerate PCR  
CC primers (AAA61714-A61715) based on conserved regions of serine  
CC proteases. The BSSP4 serine proteases and nucleotides encoding them are  
CC useful in detecting homologues, mutants and polymorphic variants in  
CC biological samples (e.g., blood, urine, brain, prostate gland and testis)  
CC as diagnostic markers for diseases associated with altered BSSP4  
CC expression levels. Such diseases include Alzheimer's disease, oedema  
CC (dropsy), cancer or inflammation of brain, prostate, testis or bone.  
CC Sequences AAA61695-A61703 and AAA61799 represent cDNAs encoding human  
CC BSSP4 variants (hBSSP4), and sequence AAA61704 represents cDNA encoding  
CC murine BSSP4 (mBSSP4). Sequences AAA61705-A61706 and AAA61731-A61732  
CC represent oligonucleotides used to construct plasmids used in the  
CC invention.

SQ Sequence 117 BP; 27 A; 38 C; 20 G; 32 T; 0 other;

Query Match 1.6%; Score 17; DB 21; Length 117;



OY 225 cgtcatggtgatgtga 241  
 |||||  
 DB 95 CGTCATGCTGATGCTGA 79

## RESULT 26

AAA61762  
 ID AAA61762 standard; DNA; 117 BP.

AAA61762;

23-OCT-2000 (first entry)

pTryptHis plasmid oligonucleotide, SEQ ID NO:32.

XX BSSP5; serine protease; human; hbSSP5; mouse; mbSSP5; brain;  
 KW diagnostic marker; antibody; transgenic animal; Alzheimer's disease;  
 KW epilepsy; cancer; inflammation; infertility; pancreatitis;  
 KW prostatic hypertrophy; plasmid construction; ss.

XX Synthetic.

XX MO200031243-A1.

XX 02-JUN-2000.

XX 19-NOV-1999; 99MO-JP06473.

XX 20-NOV-1998; 98JP-0347806.

XX (FUSO ) FUSO PHARM IND LTD.

XX Uemura H, Okui A, Kominami K, Yamaguchi N, Mitsui S;

XX WPI: 2000-400058/34.

PT Serine proteases BSSP5, useful in detecting homologs, mutants and  
 PT polymorphic variants as markers for diagnosis of e.g. Alzheimer's  
 PT disease, epilepsy, cancer and inflammation, using blood, urine,  
 PT pancreas or other tissues

XX Example 4; Page 66; 70pp; Japanese.

XX The invention relates to novel serine proteases designated BSSP5  
 CC (AA611710-B11711), and to nucleic acids encoding them (AAA61733-A61734).  
 CC The invention also relates to vectors and transformants comprising BSSP5  
 CC nucleic acids; transgenic animals in which the expression level of BSSP5  
 CC can be varied; and an mbSSP5 knockout mouse. The invention additionally  
 CC encompasses anti-BSSP5 antibodies and methods of production of such  
 CC antibodies, methods of BSSP5 detection using the antibodies, and the  
 CC use of BSSP5 proteins or fragments as diagnostic markers for certain  
 CC medical conditions, e.g., pancreatitis. A method for detecting  
 CC pancreatitis comprising measuring BSSP5 concentration in the blood or  
 CC urine, and a pancreatitis diagnostic agent containing an anti-BSSP5  
 CC antibody is also disclosed. Nucleotides encoding BSSP5 were initially  
 CC isolated in a human brain cDNA library using degenerate PCR primers  
 CC (AAA61744-A61745) based on conserved regions of serine proteases. The  
 CC BSSP5 serine proteases and nucleotides encoding them are useful in  
 CC detecting homologues, mutants and polymorphic variants in biological  
 CC samples (e.g., blood, urine, brain, prostate gland, placenta, testis,  
 CC pancreas and spleen) as diagnostic markers for conditions such as  
 CC Alzheimer's disease, epilepsy, cancer, inflammation, infertility,  
 CC pancreatitis and prostatic hypertrophy. Sequences AAA61735-A61736 and  
 CC AAA61761-A61762 represent oligonucleotides used to construct plasmids  
 CC used in the invention.

XX Sequence 117 BP; 32 A; 20 C; 38 G; 27 T; 0 other;

Query Match 1.6%; Score 17; DB 21; Length 117;  
 Best Local Similarity 100.0%; Pred. No. 87;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 225 cgtcatggtgatgtga 241  
 |||||  
 DB 23 cgtcatggtgatgtga 39

## RESULT 27

AAA61797/c  
 ID AAA61797 standard; DNA; 117 BP.

AAA61797;

23-OCT-2000 (first entry)

pTryptHis plasmid oligonucleotide, SEQ ID NO:38.

XX BSSP6; serine protease; human; hbSSP6; mouse; mbSSP6; brain;  
 KW diagnostic marker; antibody; transgenic animal; Alzheimer's disease;  
 KW epilepsy; cancer; inflammation; infertility; pancreatitis;  
 KW prostatic hypertrophy; plasmid construction; ss.

XX Synthetic.

XX MO200031257-A1.

XX 02-JUN-2000.

XX 19-NOV-1999; 99MO-JP06476.

XX 20-NOV-1998; 98JP-0347802.

XX (FUSO ) FUSO PHARM IND LTD.

XX Uemura H, Okui A, Kominami K, Yamaguchi N, Mitsui S;

XX WPI: 2000-400067/34.

PT Serine protease BSSP6, useful in detecting homologs, mutants and  
 PT polymorphic variants as markers for diagnosis of Alzheimer's disease,  
 PT epilepsy, cancer, inflammation, infertility and prostate hypertrophy,  
 PT using blood or other tissues

XX Example 4; Page 90; 94pp; Japanese.

XX The invention relates to novel serine proteases designated BSSP6  
 CC (AA611712-B11714), and to nucleic acids encoding them (AAA61763-A61765).  
 CC The invention also relates to vectors and transformants comprising BSSP6  
 CC nucleic acids; transgenic animals in which the expression level of BSSP6  
 CC can be varied; and an mbSSP6 knockout mouse. The invention additionally  
 CC encompasses anti-BSSP6 antibodies and methods of production of such  
 CC antibodies, methods of BSSP6 detection using the antibodies, and the  
 CC use of BSSP6 proteins or fragments as diagnostic markers for certain  
 CC medical conditions. Nucleotides encoding BSSP6 were initially  
 CC isolated in a human brain cDNA library using degenerate PCR primers  
 CC (AAA61795-A61796) based on conserved regions of serine proteases. The  
 CC BSSP6 serine proteases and nucleotides encoding them are useful in  
 CC detecting homologues, mutants and polymorphic variants in biological  
 CC samples (e.g., blood, urine, brain, prostate gland, placenta, testis,  
 CC pancreas and spleen) as diagnostic markers for conditions such as Alzheimer's  
 CC disease, epilepsy, cancer, inflammation, infertility and prostatic  
 CC hypertrophy. Sequences AAA61766-A61767 and AAA61797-A61798 represent  
 CC oligonucleotides used to construct plasmids used in the invention.

XX Sequence 117 BP; 27 A; 38 C; 20 G; 32 T; 0 other;

Query Match 1.6%; Score 17; DB 21; Length 117;  
 Best Local Similarity 100.0%; Pred. No. 87;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 225 cgtcatggtgatgtga 241  
 |||||  
 DB 95 CGTCATGCTGATGCTGA 79



RESULT 28  
ID AAA61798 standard; DNA: 117 BP.  
AC AAA61798;  
XX  
XX 23-OCT-2000 (first entry)  
DE pTrypHis plasmid oligonucleotide, SEQ ID NO:39.  
XX  
XX BSSP6; serine protease; human; hbSSP6; mouse; mSSP6; brain;  
XX diagnostic marker; antibody; transgenic animal; Alzheimer's disease;  
XX epilepsy; cancer; inflammation; infertility; pancreatitis;  
XX prostatic hypertrophy; plasmid construction; ss.  
XX  
XX Synthetic.  
XX  
XX WO200031257-A1.  
XX  
XX 02-JUN-2000.  
XX  
XX 19-NOV-1999; 99WO-JP06476.  
XX  
XX 20-NOV-1998; 98JP-0347802.  
XX  
XX (FUSO ) FUSO PHARM IND LTD.  
XX  
XX Uemura H, Okui A, Komlinami K, Yamaguchi N, Mitsui S;  
XX  
XX WPI: 2000-40067/34.  
XX  
XX Serine protease BSSP6, useful in detecting homologs, mutants and  
XX polymorphic variants as markers for diagnosis of Alzheimer's disease,  
XX epilepsy, cancer, inflammation, infertility and prostate hypertrophy,  
XX using blood or other tissues -  
XX  
XX Example 4; Page 90; 94pp; Japanese.  
XX  
XX The invention relates to novel serine proteases designated BSSP6  
XX (AA61712-B11714), and to nucleic acids encoding them (AA61763-A61765).  
XX The invention also relates to vectors and transformants comprising BSSP6  
XX nucleic acids; transgenic animals in which the expression level of BSSP6  
XX can be varied; and an mSSP6 knockout mouse. The invention additionally  
XX encompasses anti-BSSP6 antibodies and methods of production of such  
XX antibodies, methods of BSSP6 detection using the antibodies, and the  
XX use of BSSP6 proteins or fragments as diagnostic markers for certain  
XX medical conditions. Nucleotides encoding BSSP6 were initially  
XX isolated in a human brain cDNA library using degenerate PCR primers  
XX (AA61795-A61796) based on conserved regions of serine proteases. The  
XX BSSP6 serine proteases and nucleotides encoding them are useful in  
XX detecting homologues, mutants and polymorphic variants in biological  
XX samples (e.g., blood, urine, brain, prostate gland, placenta, testis  
XX and spleen) as diagnostic markers for conditions such as Alzheimer's  
XX disease, epilepsy, cancer, inflammation, infertility and prostatic  
XX hypertrophy. Sequences AAA61766-A61767 and AAA61797-A61798 represent  
XX oligonucleotides used to construct plasmids used in the invention.  
XX  
XX Sequence 117 BP; 32 A; 20 C; 38 G; 27 T; 0 other;

RESULT 30  
ID AAA62803 standard; DNA: 117 BP.  
AC AAA62803;  
XX  
XX 20-OCT-2000 (first entry)  
DE Oligonucleotide #1 used in plasmid pTrypHis construction.  
XX  
XX Expression vector; active target protein; easy isolation; pTrypHis; ss.  
XX  
XX Synthetic.  
XX  
XX WO200031284-A1.  
XX  
XX 02-JUN-2000.  
XX  
XX 19-NOV-1999; 99WO-JP06474.  
XX  
XX 20-NOV-1998; 98JP-0331515.  
XX  
XX (FUSO ) FUSO PHARM IND LTD.  
XX  
XX Uemura H, Okui A, Komlinami K, Yamaguchi N, Mitsui S;  
XX  
XX WPI: 2000-411958/35.  
XX  
XX protein expression vectors containing specific sequences with cloning  
XX sites into which nucleic acids insert, for isolation of target proteins  
XX without excess amino acids -  
XX  
XX Example 1; Page 33; 44pp; Japanese.  
XX  
XX The invention relates to a protein expression vector from which mature  
XX and active target proteins can be easily isolated. N and C-terminal amino  
XX acids are kept to a minimum to ensure that easy isolation from the vector  
XX is possible. The protein expression vector comprises a secretory signal  
XX nucleic acid sequence and, in the 3'-downstream side, a tag nucleic acid  
XX sequence, a scissile nucleic acid sequence and a cloning site into which  
XX a nucleic acid sequence encoding a target protein can be inserted. Also  
XX included in the invention are a host cell transformed by the vector, a  
XX process for producing a target protein using the vector or host cell, a  
XX recombinant fusion protein containing the amino acid sequence of the  
XX target protein, and a process for the production of mature and active target  
XX expression vector is used for the production of mature and active target  
XX protein. The present sequence represents an oligonucleotide use in the  
XX construction of the plasmid pTrypHis, for use in the invention.  
XX  
XX Sequence 117 BP; 27 A; 38 C; 20 G; 32 T; 0 other;

Query Match 1.6%; Score 17; DB 21; Length 117;  
Best Local Similarity 100.0%; Pred. NO. 87;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 225 cgtcatggtgatgtgta 241  
|||||  
DB 95 cgtcatggtgatgtgta 79  
|||||

RESULT 30  
ID AAA62804 standard; DNA: 117 BP.  
AC AAA62804;  
XX  
XX 20-OCT-2000 (first entry)  
DE Oligonucleotide #2 used in plasmid pTrypHis construction.  
XX  
XX Expression vector; active target protein; easy isolation; pTrypHis; ss.  
XX  
XX Synthetic.  
XX

PM WO200031284-A1.

XX 02-JUN-2000.

PF 19-NOV-1999; 99WO-JP06474.

PR 20-NOV-1998; 98JP-0331515.

PA (FUSO ) FUSO PHARM IND LTD.

PI Uemura H, Okui A, Komimami K, Yamaguchi N, Mitsui S;

DR WPI: 2000-411958/35.

XX Protein expression vectors containing specific sequences with cloning  
PT sites into which nucleic acids insert, for isolation of target proteins  
PT without excess amino acids -

PS Example 1; Page 33-34; 44pp; Japanese.

CC The invention relates to a protein expression vector from which mature  
CC and active target proteins can be easily isolated. N and C-terminal amino  
CC acids are kept to a minimum to ensure that easy isolation from the vector  
CC is possible. The protein expression vector comprises a secretory signal  
CC nucleic acid sequence and, in the 3'-downstream side, a tag nucleic acid  
CC sequence, a scissile nucleic acid sequence and a cloning site into which  
CC a nucleic acid sequence encoding a target protein can be inserted. Also  
CC included in the invention are a host cell transformed by the vector, a  
CC process for producing a target protein containing the amino acid sequence of the  
CC recombinant fusion protein containing the amino acid sequence of the  
CC target protein, and a process for producing a target protein. The  
CC expression vector is used for the production of mature and active target  
CC protein. The present sequence represents an oligonucleotide use in the  
CC construction of the plasmid pTryptis, for use in the invention.

SO Sequence 117 BP; 32 A; 20 C; 38 G; 27 T; 0 other;

Query Match 1.6%; Score 17; DB 21; Length 117;

Best Local Similarity 100.0%; Pred. No. 87;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 225 cgtcatggtatggtgga 241

DB 23 cgtcatggtatggtgga 39

RESULT 31

AAA62817/C

ID AAA62817 standard; DNA; 120 BP.

XX AAA62817;

DT 20-OCT-2000 (first entry)

DE DNA encoding active human neurosin.

KW Expression vector; active target protein; easy isolation; neurosin;

KW human; ds.

OS Homo sapiens.

PM WO200031284-A1.

PD 02-JUN-2000.

PF 19-NOV-1999; 99WO-JP06474.

PR 20-NOV-1998; 98JP-0331515.

PA (FUSO ) FUSO PHARM IND LTD.

PI Uemura H, Okui A, Komimami K, Yamaguchi N, Mitsui S;

XX WPI: 2000-411958/35.

DR P-PSDB; AAB03864.

XX Protein expression vectors containing specific sequences with cloning  
PT sites into which nucleic acids insert, for isolation of target proteins  
PT without excess amino acids -

PS Example 3; Page 42; 44pp; Japanese.

CC The invention relates to a protein expression vector from which mature  
CC and active target proteins can be easily isolated. N and C-terminal amino  
CC acids are kept to a minimum to ensure that easy isolation from the vector  
CC is possible. The protein expression vector comprises a secretory signal  
CC nucleic acid sequence and, in the 3'-downstream side, a tag nucleic acid  
CC sequence, a scissile nucleic acid sequence and a cloning site into which  
CC a nucleic acid sequence encoding a target protein can be inserted. Also  
CC included in the invention are a host cell transformed by the vector, a  
CC process for producing a target protein using the amino acid sequence of the  
CC recombinant fusion protein containing the amino acid sequence of the  
CC target protein, and a process for producing a target protein. The  
CC expression vector is used for the production of mature and active target  
CC protein. The present sequence represents DNA encoding active human  
CC neurosin. Neurosin is used as the target protein in an example of the  
CC vector of the invention.

SO Sequence 120 BP; 28 A; 33 C; 25 G; 34 T; 0 other;

Query Match 1.6%; Score 17; DB 21; Length 120;

Best Local Similarity 100.0%; Pred. No. 87;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 225 cgtcatggtatggtgga 241

DB 109 cgtcatggtatggtgga 93

RESULT 32

AAA62812/C

ID AAA62812 standard; DNA; 135 BP.

XX AAA62812;

DT 20-OCT-2000 (first entry)

DE DNA encoding active human neurosin.

KW Expression vector; active target protein; easy isolation; neurosin;

KW human; ds.

OS Homo sapiens.

PM WO200031284-A1.

PD 02-JUN-2000.

PF 19-NOV-1999; 99WO-JP06474.

PR 20-NOV-1998; 98JP-0331515.

PA (FUSO ) FUSO PHARM IND LTD.

PI Uemura H, Okui A, Komimami K, Yamaguchi N, Mitsui S;

DR WPI: 2000-411958/35.

DR P-PSDB; AAB03863.

XX Protein expression vectors containing specific sequences with cloning  
PT sites into which nucleic acids insert, for isolation of target proteins  
PT without excess amino acids -

PS Example 2; Page 41; 44pp; Japanese.

XX The invention relates to a protein expression vector from which mature  
CC and active target proteins can be easily isolated. N and C-terminal amino  
CC acids are kept to a minimum to ensure that easy isolation from the vector  
CC is possible. The protein expression vector comprises a secretory signal  
CC nucleic acid sequence and, in the 3'-downstream side, a 799 nucleic acid  
CC sequence, a scissile nucleic acid sequence and a cloning site into which  
CC a nucleic acid sequence encoding a target protein can be inserted. Also  
CC included in the invention are a host cell transformed by the vector, a  
CC process for producing a target protein using the vector or host cell, a  
CC recombinant fusion protein containing the amino acid sequence of the  
CC target protein, and a process for producing a target protein. The  
CC expression vector is used for the production of mature and active target  
CC protein. The present sequence represents DNA encoding active human  
CC neurosin. Neurosin is used as the target protein in an example of the  
CC vector of the invention.

SQ Sequence 135 BP; 29 A; 45 C; 38 G; 23 T; 0 other;

Query Match 1.6%; Score 17; DB 21; Length 135;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 225 cgtatgtatgtatgtatga 241  
|||||  
DB 124 CGTCATGTCATGTCATGA 108

RESULT 33

AAAA2073/C  
ID AAAA2073 standard; CDNA; 178 BP.

AC AAAA2073;

XX 21-AUG-2000 (first entry)

DE Human secreted expressed sequence tag SEQ ID NO:813.

XX Human; mouse; xenopus; rat; secreted expressed sequence tag; SEST;  
XX expressed sequence tag; EST; probe; chemotactic; proliferative;  
XX immunomodulatory; haematopoietic; chemokine; analgesic; haemostatic;  
XX thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;  
XX antiviral; antidiabetic; antisthmatic; vulnery; antiparkinsonian;  
XX antitumor; osteoprotective; neuroprotective; nootropic; antipsoriatic;  
XX cerebroprotective; anticonvulsant; antidepressant; gene therapy;  
XX vaccine; autoimmune disorder; multiple sclerosis; allergic condition;  
XX insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;  
XX lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;  
XX central nervous system disorder; Alzheimer's disease; stroke;  
XX Parkinson's disease; Huntington's disease; coagulation disorder;  
XX haemophilia; thrombosis; inflammatory disorder; Crohn's disease;  
XX tumour; infection; depression; psoriasis; ss.

XX Homo sapiens.

OS WO200021990-A1.

XX 20-APR-2000.

XX 15-OCT-1999; 99WO-US24205.

XX 15-OCT-1996; 98US-0104435.

XX (GEM ) GENETICS INST INC.

XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;

XX Merberg D, Treacy M;

XX WPI; 2000-317937/27.

XX Isolated polynucleotides, and encoded proteins, comprising secreted  
XX expressed sequence tags (SESTs), useful for treating various disorders

PT such as autoimmune, infectious, and central nervous system disorders -  
XX  
XX Claim 1; Page 339; 618pp; English.

XX AAA41261 to AAA43419 represent specifically claimed secreted expressed  
XX sequence tags (SESTs), isolated from human, mouse, xenopus and rat  
XX tissue sources. The SESTs can have a range of activities depending on  
XX the tissues they were isolated from. The activities include:  
XX chemotactic; proliferative; immunomodulatory; haematopoietic;  
XX thrombolytic; antiinflammatory;  
XX cytoskeletal; analgesic; haemostatic; thrombolytic; antidiabetic;  
XX antisthmatic; vulnery; antitumor; osteoprotective; neuroprotective;  
XX nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;  
XX anticonvulsant; and antidepressant. The SESTs can be used for gene  
XX therapy and in vaccines. The SESTs are useful as probes for the  
XX identification and isolation of full-length cDNAs and genomic DNA  
XX molecules which correspond to the SESTs. Proteins encoded by the SESTs  
XX are useful in assays for determining biological activity and raising  
XX antibodies. They may be useful for treatment of autoimmune disorders  
XX (multiple sclerosis, insulin dependent diabetes), allergic conditions  
XX (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,  
XX osteoporosis, osteoarthritis, central nervous system disorders  
XX (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation  
XX disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's  
XX disease), tumours, bacterial, fungal or viral infections, depression and  
XX psoriasis. AAA43420 to AAA43425 represent linker variants which are given  
XX in the exemplification of the present invention.

SQ Sequence 178 BP; 42 A; 36 C; 30 G; 70 T; 0 other;

Query Match 1.6%; Score 17; DB 21; Length 178;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 737 agttgcagatgacatg 753  
|||||

DB 174 AGTTTGCAGATGACATG 158

RESULT 34

AXA11731  
ID AXA11731 standard; DNA; 227 BP.

AC AXA11731;

XX 30-MAR-1999 (first entry)

DE Human biallelic polymorphic DNA fragment ESTD-VB12b.

XX polymorphism; biallelic; human; forensic; paternity testing; disease;  
XX detection; phenotypic typing; characteristic; infection; hereditary;  
XX autoimmune disease; cancer; inflammation; drug; therapy; medication;  
XX treatment; marker; ss.

XX Homo sapiens.

OS WO9820165-A2.

XX 14-MAY-1998.

XX 05-NOV-1997; 97WO-US20313.

XX 06-NOV-1996; 96US-0030455.

XX (WHED ) WHITEHEAD INST BIOMEDICAL RES.

XX Hudson T, Lander ES, Wang D;

XX WPI; 1998-286974/25.

XX New isolated nucleic acid segments from the human genome - used for  
XX determining polymorphic forms for use in e.g. forensics, paternity

PT testing or phenotypic typing for disease

XX Claim 1; Page 187; 310pp; English.

CC AAX10269-X12937 are human DNA fragments which contain biallelic  
CC polymorphic markers which have been isolated using the primers  
CC represented in AAX09121-X10268. The base occupying the polymorphic site  
CC is indicated by the appropriate IUPAC-IUB ambiguity code. These fragments  
CC can be used in methods for determining polymorphic forms in an individual  
CC for use in e.g. forensics, paternity testing or for phenotypic typing for  
CC diseases such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan  
CC syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease,  
CC familial hypercholesterolemia, polycystic kidney disease, hereditary  
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary  
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular  
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases.

SQ Sequence 227 BP; 65 A; 55 C; 52 G; 54 T; 1 other;

Query Match 1.6%; Score 17; DB 19; Length 227;

Best Local Similarity 100.0%; Pred. No. 87;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 cagggtcgtgttttc 90  
DB 45 cagggtcgtgttttc 61

RESULT 35

AAX11732 standard; DNA; 227 BP.

XX AAX11732;

XX 30-MAR-1999 (first entry)

DE Human biallelic polymorphic DNA fragment ESTD-VB12a.

XX Polymorphism; biallelic; human; forensics; paternity testing; disease;  
KW detection; phenotypic typing; characteristic; infection; hereditary;  
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;  
XX treatment; marker; ss.

OS Homo sapiens.

XX W09820165-A2.

PN 14-MAY-1998.

PD 05-NOV-1997; 97WO-US20313.

PR 06-NOV-1996; 96US-0030455.

XX (WHEED ) WHITEHEAD INST BIOMEDICAL RES.

XX Hudson T, Lander ES, Wang D;

XX WPI; 1998-286974/25.

PT New isolated nucleic acid segments from the human genome - used for  
PT determining polymorphic forms for use in e.g. forensics, paternity  
PT testing or phenotypic typing for disease

PS Claim 1; Page 187; 310pp; English.

CC AAX10269-X12937 are human DNA fragments which contain biallelic

CC polymorphic markers which have been isolated using the primers  
CC represented in AAX09121-X10268. The base occupying the polymorphic site  
CC is indicated by the appropriate IUPAC-IUB ambiguity code. These fragments  
CC can be used in methods for determining polymorphic forms in an individual  
CC for use in e.g. forensics, paternity testing or for phenotypic typing for  
CC diseases such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan  
CC syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease,  
CC familial hypercholesterolemia, polycystic kidney disease, hereditary  
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary  
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular  
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases.

SQ Sequence 227 BP; 64 A; 56 C; 52 G; 54 T; 1 other;

Query Match 1.6%; Score 17; DB 19; Length 227;

Best Local Similarity 100.0%; Pred. No. 87;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 cagggtcgtgttttc 90  
DB 45 cagggtcgtgttttc 61

RESULT 36

AAX11730 standard; DNA; 227 BP.

XX AAX11730;

XX 30-MAR-1999 (first entry)

DE Human biallelic polymorphic DNA fragment ESTD-VB12.

XX Polymorphism; biallelic; human; forensics; paternity testing; disease;  
KW detection; phenotypic typing; characteristic; infection; hereditary;  
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;  
XX treatment; marker; ss.

OS Homo sapiens.

XX W09820165-A2.

PN 14-MAY-1998.

PD 05-NOV-1997; 97WO-US20313.

PR 06-NOV-1996; 96US-0030455.

XX (WHEED ) WHITEHEAD INST BIOMEDICAL RES.

XX Hudson T, Lander ES, Wang D;

XX WPI; 1998-286974/25.

PT New isolated nucleic acid segments from the human genome - used for  
PT determining polymorphic forms for use in e.g. forensics, paternity  
PT testing or phenotypic typing for disease

PS Claim 1; Page 187; 310pp; English.

CC AAX10269-X12937 are human DNA fragments which contain biallelic  
CC polymorphic markers which have been isolated using the primers  
CC represented in AAX09121-X10268. The base occupying the polymorphic site  
CC is indicated by the appropriate IUPAC-IUB ambiguity code. These fragments



CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 322 BP; 84 A; 53 C; 55 G; 119 T; 11 other;

Query Match 1.6%; Score 17; DB 16; Length 322;  
Best Local Similarity 100.0%; Pred. No. 86;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 683 aagactctgaacattc 699  
|||||  
DB 112 AAGACTCTGACCAATTC 96

RESULT 39  
AAC03135  
ID AAC03135 standard; cDNA; 364 BP.

XX AAC03135;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 3133.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI; 2000-500381/45.

DR P-PSDB; AAG03129.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX  
XX Claim 1; SEQ ID 3133; 71pp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. An ORF has been identified within the  
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
CC derived from 30 different tissues. EST sequences usually correspond  
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
CC well suited for isolating cDNA sequences derived from the 5' ends of  
CC mRNAs and even in those cases where longer cDNA sequences have been  
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
CC gene therapy and chromosome mapping procedures. They are used to obtain  
CC upstream regulatory sequences and to design expression and secretion  
CC vectors.

SQ Sequence 364 BP; 92 A; 83 C; 86 G; 103 T; 0 other;

Query Match 1.6%; Score 17; DB 21; Length 364;  
Best Local Similarity 100.0%; Pred. No. 86;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 931 cttagataatgacgc 947  
|||||  
DB 139 cttagataatgacgc 155

RESULT 40  
AAA3106/c  
ID AAA3106 standard; DNA; 373 BP.

XX AAA3106;

DT 05-JUL-2000 (first entry)

DE Plant microsatellite marker #67.

KW Plant microsatellite sequence; core repeat sequence; detection; probe;  
KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;  
KW variety identification; genetic variability evaluation; primer; ss.

XX Eucalyptus grandis.

PN WO9967421-A1.

PD 29-DEC-1999.

PF 25-JUN-1999; 99WO-NZ00092.

PR 25-JUN-1998; 98US-0105307.

PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.

PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.

PI Havukkala IJ, Blocksberg LN, Glenn M;

DR WPI; 2000-116958/10.

PT New plant microsatellite markers and associated flanking species for  
PT the detection of polymorphic genetic markers -  
XX  
XX Claim 1; Page 96; 392pp; English.

CC Sequences AAA31040-A32093 represent novel plant microsatellite sequences  
CC and associated flanking species. The sequences comprise a central core  
CC repeat sequence, especially selected from the sequences AAA32094-A32096  
CC with left and right flanking sequences. The polynucleotide sequences  
CC can be used in the detection of DNA polymorphisms, in genome mapping,  
CC in physical mapping, in positional cloning of genes, in variety  
CC identification and in evaluation of genetic variability within and  
CC between plant tissues, populations, cultivars, species and species  
CC groups. They may also be used to design hybridization probes for  
CC primers for microsatellite-primed PCR. Microsatellite markers are  
CC useful to locate specific economically useful genes in plant genomes.  
SQ Sequence 373 BP; 97 A; 85 C; 104 G; 87 T; 0 other;

Query Match 1.6%; Score 17; DB 21; Length 373;  
Best Local Similarity 100.0%; Pred. No. 86;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 617 atcatctcctgattta 633  
|||||  
DB 301 ATCATCTCTGATTTA 285

RESULT 41  
AAC19113/c

ID AAC19113 standard; cDNA; 425 BP.

XX AAC19113;

```

XX 06-OCT-2000 (first entry)
XX Human secreted protein 5' EST, SEQ ID NO: 23188.
XX Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.
XX Homo sapiens.
XX EPI033401-A2.
XX 06-SEP-2000.
XX 21-FEB-2000; 2000EP-0200610.
XX 26-FEB-1999; 99US-0122487.
XX (GENSET) GENSET.
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 1; SEQ ID 23188; 71pp + CD-ROM; English.
XX The present sequence is one of a large number of 5' ESTs derived from
XX cDNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors.
XX Sequence 425 BP; 132 A; 84 C; 85 G; 124 T; 0 other;
XX
XX Query Match 1.6%; Score 17; DB 21; Length 425;
XX Best local Similarity 100.0%; Pred. No. 86;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 476 ttcaagaactcgtt 492
XX |||||||||||||||
XX 174 TTCAGAAACTCTGTT 158
XX
XX RESULT 42
XX AAC09333/c
XX ID AAC09333 standard; cDNA; 446 BP.
XX AAC09333;
XX 06-OCT-2000 (first entry)
XX Human secreted protein 5' EST, SEQ ID NO: 13408.
XX Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.
XX Homo sapiens.
XX EPI033401-A2.

```

```

XX 06-SEP-2000.
XX 21-FEB-2000; 2000EP-0200610.
XX 26-FEB-1999; 99US-0122487.
XX (GENSET) GENSET.
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 1; SEQ ID 13408; 71pp + CD-ROM; English.
XX The present sequence is one of a large number of 5' ESTs derived from
XX cDNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors.
XX Sequence 446 BP; 134 A; 76 C; 70 G; 162 T; 4 other;
XX
XX Query Match 1.6%; Score 17; DB 21; Length 446;
XX Best local Similarity 100.0%; Pred. No. 86;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 683 aagacttgaacattc 699
XX |||||||||||||||
XX 168 AAGACTCTGACACATTC 152
XX
XX RESULT 43
XX AAX23298
XX ID AAX23298 standard; DNA; 701 BP.
XX AAX23298;
XX 11-JUN-1999 (first entry)
XX Human TRYI trypsinogen variant cDNA.
XX TRYI; trypsinogen; autocatalysis; cleavage; zymogenic precursor;
XX protease; peptide hormone; therapeutic protein; treatment; coagulation;
XX ss.
XX Homo sapiens.
XX OS
XX WO910503-A1.
XX 04-MAR-1999.
XX 12-AUG-1998; 98MO-EP05094.
XX 15-OCT-1997; 97EP-0117816.
XX 22-AUG-1997; 97EP-0114513.
XX (HOFF) ROCHE DIAGNOSTICS GMBH.

```





PT diagnosing, preventing, treating or ameliorating medical conditions and  
 PT used as food additives or preservatives -

PS Claim 1; Page 351; 410pp; English.

CC The invention relate to the isolation of genes AAC93310-C93354 encoding  
 CC 45 human secreted proteins AAB51380-B51423. The genes can be used to  
 CC generate fusion proteins by linking to the gene for the human  
 CC immunoglobulin G Fc portion (AAC93301) for increasing the stability of  
 CC the fusion protein as compared to the human protein only. The genes and  
 CC proteins are useful for preventing, ameliorating or treating medical  
 CC conditions, e.g. by protein or gene therapy. The genes are isolated  
 CC from a range of human tissues disclosed in the specification. The  
 CC nucleic acids, proteins, antibodies and (ant)agonists are useful in  
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, diabetes mellitus, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)  
 CC wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.

SQ Sequence 936 BP; 278 A; 178 C; 208 G; 269 T; 3 other;

Query Match 1.68; Score 17; DB 21; Length 936;  
 Best Local Similarity 100.0%; Pred. No. 85;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 512 attctcgtagtagaac 528  
 |||||||||  
 Db 753 attctcgtagtagaac 769

#### RESULT 46

AA054636/C  
 ID AA054636 standard; DNA; 1104 BP.

XX  
 AC AA054636;

DT 22-JUN-1994 (first entry)

DE Guinea pig PH-30, 30 alpha subunit.

KW PH-20; PH-30; contraceptive; fertilisation; sperm surface protein;

KW vaccine; sperm-egg fusion; ds.

XX Cavia cobaya.

OS NO9325233-A.

PN 23-DEC-1993.

PD 10-JUN-1993; 93WO-US05640.

PR 12-JUN-1992; 92US-0897883.

XX (UYCO-) UNIV CONNECTICUT.

PI Myles DG, Primakoff P;

DR WPI; 1994-007200/01.

XX P-PSDB; AAR52633.

PT Contraceptive vaccine for reducing sperm-egg fusion - comprises  
 PT peptide from sperm surface protein which stimulates antibody  
 prodn.

XX Example 5; Fig 6; 79pp; English.

CC Sperm surface proteins or peptides stimulate an immune response to  
 CC produce antibodies which block sperm-egg fusion and provide  
 CC contraception. Pref. sperm surface proteins are the PH-20 and PH-30  
 CC sperm surface proteins.

SQ Sequence 1104 BP; 345 A; 240 C; 251 G; 268 T; 0 other;

Query Match 1.68; Score 17; DB 15; Length 1104;  
 Best Local Similarity 100.0%; Pred. No. 85;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 324 tttgtcaattcattc 340  
 |||||||||  
 Db 526 TGTGTCAANTCATTTC 510

#### RESULT 47

AA15908/C  
 ID AA15908 standard; cDNA; 1619 BP.

XX  
 AC AA15908;

DT 12-JUN-2000 (first entry)

DE Human protein clone HP10195 full length coding sequence.

XX Human protein; hydrophobic domain; nutritional source; hematopoiesis;  
 KW cytokine production; cell proliferation; cell differentiation;  
 KW immune deficiency; infectious disease; autoimmune disorder; asthma;  
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;  
 KW allergic reaction; osteoporosis; osteoarthritis; periodontal disease;  
 KW nervous system disorder; Alzheimer's disease; Parkinson's disease;  
 KW Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury;  
 KW systemic cytokine damage; tissue differentiation; contraceptive; stroke;  
 KW coagulation disorder; myocardial infarction; inflammatory condition;  
 KW septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour;  
 KW nephritis; therapy; ss.

XX OS Homo sapiens.

PN WO200005367-A2.

PD 03-FEB-2000.

PE 22-JUL-1999; 99WO-JP03929.

PR 24-JUL-1998; 98JP-0208820.

PR 07-AUG-1998; 98JP-0224105.

PR 25-AUG-1998; 98JP-0238116.

PR 09-SEP-1998; 98JP-0254736.

PR 29-SEP-1998; 98JP-0275505.

XX (SAGA ) SAGAMI CHEM RES CENT.

PA (PROT-) PROTEGENE INC.

PI Kato S, Kimura T;

DR WPI; 2000-182694/16.

XX P-PSDB; AAT94850.

PT Novel human proteins having hydrophobic domains useful for treating  
 PT osteoporosis, Alzheimer's disease, Parkinson's disease, asthma,  
 PT multiple sclerosis, rheumatoid arthritis, cancer, anaemia, and stroke -

Claim 4; Page 192-194; 351pp; English.

XX This sequence encodes a human protein of the invention, which has  
 CC hydrophobic domains. The DNA sequences can be used as a probe or as a  
 CC genetic marker. The protein can also be used as a marker, and to identify  
 CC potential genetic disorders. The DNA and protein can also be used as  
 CC nutritional sources or supplements. The protein exhibits cytokine, cell  
 CC proliferation, cell differentiation activities and induces production of

QY	683	aagactctgaacaattc	699
Db	1338	AAGACTCTGAACAATTc	1352

AC	AAC53614;
XX	
DT	18-OCT-2000 (first entry)
XX	

XX  
KW  
KW  
KW  
KW  
XX

PN EP1033405-A2  
xx

PF 25-FEB-2000; 2000EP-0301439.  
 VV

PR 09-MAR-1999; 99US-0123548.

01-APR-1999; 99US-0127462.

PR 16-APR-1999; 99US-0129845.

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

[illegible]

PR	14-OCT-1999;	990S-0159330.
PR	14-OCT-1999;	990S-0159331.
PR	14-OCT-1999;	990S-0159637.
PR	14-OCT-1999;	990S-0159638.
PR	18-OCT-1999;	990S-0159584.
PR	21-OCT-1999;	990S-0160741.
PR	21-OCT-1999;	990S-0160767.
PR	21-OCT-1999;	990S-0160768.
PR	21-OCT-1999;	990S-0160770.
PR	21-OCT-1999;	990S-0160814.
PR	21-OCT-1999;	990S-0160815.
PR	22-OCT-1999;	990S-0160980.
PR	22-OCT-1999;	990S-0160981.
PR	22-OCT-1999;	990S-0160989.
PR	25-OCT-1999;	990S-0161404.
PR	25-OCT-1999;	990S-0161405.
PR	25-OCT-1999;	990S-0161406.
PR	26-OCT-1999;	990S-0161359.
PR	26-OCT-1999;	990S-0161360.
PR	26-OCT-1999;	990S-0161361.
PR	28-OCT-1999;	990S-0161920.
PR	28-OCT-1999;	990S-0161992.
PR	28-OCT-1999;	990S-0161993.
PR	29-OCT-1999;	990S-0162142.

Query Match	1.66;	Score 17;	DB 21;	Length 1676;
Best Local Similarity	100.0%;	Pred. No. 84;		
Matches 17;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Oy	1026 gataacaccagaagatc 1042			
DB	846 GACAAACCCAGAGATTC 830			
RESULT 49				
AD	AAc54823/c			
ID	AAc54823 standard; DNA; 1676 BP.			
XX				
AC	AAc54823;			
XX				
DT	18-OCT-2000 (first entry)			
XX				
DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 79218.			
XX				
KM	Hybridisation assay; genetic mapping; gene expression control;			
KW	protein identification; signal transduction pathway;			
XX	metabolic pathway; promoter; termination sequence; ss.			
OS	Arabidopsis thaliana.			
XX				
PN	EP1033405-A2.			
XX				
PD	06-SEP-2000.			
XX				
PF	25-FEB-2000; 2000EP-0301439.			
XX				
PR	25-FEB-1999; 990S-012825.			
PR	05-MAR-1999; 990S-0123180.			
PR	09-MAR-1999; 990S-0123548.			
PR	23-MAR-1999; 990S-0125788.			
PR	25-MAR-1999; 990S-0126264.			
PR	29-MAR-1999; 990S-0126785.			
PR	01-APR-1999; 990S-0127462.			
PR	06-APR-1999; 990S-0128234.			
PR	08-APR-1999; 990S-0128714.			
PR	16-APR-1999; 990S-0129845.			
PR	19-APR-1999; 990S-0130077.			
PR	21-APR-1999; 990S-0130449.			
PR	23-APR-1999; 990S-0130510.			
PR	28-APR-1999; 990S-0130891.			
PR	30-APR-1999; 990S-0131449.			
PR	30-APR-1999; 990S-0132048.			
PR	30-APR-1999; 990S-0132407.			



PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 22-OCT-1999; 99US-0161404.  
 PR 23-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 25-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161922.  
 PR 28-OCT-1999; 99US-0161923.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 1.6%; Score 17; DB 21; Length 1676;  
 Best Local Similarity 100.0%; Pred. No. 84;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1026 gacaacacagaagattc 1042  
 |||||  
 DB 846 GACAAACCAAGATTC 830

RESULT 50  
 AAV71287/c

ID AAV71287 standard; DNA; 1723 BP.

XX AAV71287;

DT 12-AUG-1999 (first entry)

XX Human vesicular binding protein nucleotide sequence.

XX VSBP; fibroblast; Aplysia; VAMP; vesicle associated membrane protein;

KW synaptobrevin binding protein; cancer; inflammation; ss.

XX Homo sapiens.

OS Key Location/Qualifiers

FT CDS 568..1278

FT /tag= a

FT /product= "MVBP"

XX MO9851797-A1.

PD 19-NOV-1998.

XX 15-MAY-1998; 98WO-US10225.

XX 15-MAY-1997; 97US-0857213.

XX (INCY-) INCYTE PHARM INC.

PI Bandman O, Hawkins PR, Murry LE;

XX WPI, 1999-045229/04.

DR P-PSDB; AAM85813.

XX Human vesicle binding protein - useful for the diagnosis, treatment

PT and prevention of cell proliferation disorders and inflammation

XX Claim 5; Fig 1A-D; 58pp; English.

XX The present sequence represents a nucleic acid encoding a purified human  
 CC vesicle binding protein (designated MVBP). This was obtained by standard  
 CC cloning and screening procedure, where it was first isolated in incyte  
 CC clone 148415 from a normal fibroblast library (PIBRNGT01). A comparison  
 CC of the MVBP with the Aplysia vesicle-associated membrane protein  
 CC (VAMP)/synaptobrevin binding protein showed that they shared 44%

CC identity. MVBP and its products can be used to diagnose, treat or prevent  
 CC disorders of cell proliferation (e.g. cancer) or inflammation. Sequences  
 CC complementary to MVBP can be used to detect MVBP in samples. Ads are used  
 CC to diagnose conditions associated with MVBP or in assays to monitor  
 CC patients being treated with MVBP or agonists/antagonists.

XX Sequence 1723 BP; 548 A; 356 C; 325 G; 468 T; 26 other;

Query Match 1.6%; Score 17; DB 20; Length 1723;

Best Local Similarity 100.0%; Pred. No. 84;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 683 aagactctgacaattc 699  
 |||||  
 DB 1601 AAGACTCTGACAAATTC 1585

Search completed: August 19, 2001, 00:24:14

Job time: 9000 sec

Mon Aug 20 10:21:31 2001

us-09-284-320-31.01110.rng

---

Page 38



```
117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
```

---

```
190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



SUMMARIES				Result	
No.	Score	Query Match Length	ID	Description	
1	652	62.1	BG30418	BG30418 602559169	68
2	591	56.3	BG288867	BG288867 602383808	311
3	567	54.0	BE897106	BE897106 601439663	29.6
4	529	50.4	BG403007	BG403007 602418818	458
5	516	49.1	BF060652	BF060652 7159404.1	14
6	489	47.5	BG501143	BG501143 602546382	753
7	480	45.7	BF196461	BF196461 h786607.x	302
8	478	45.0	BF132377	BF132377 n611a03.x	28.7
9	452	43.0	AM024459	AM024459 wu6106.x	452
10	450	42.7	AM003884	AM003884 ws61104.x	189
11	448	42.7	AM003884	AM003884 EST370970	159
12	440	41.9	AA393599	AA393599 zt72e08.r	14
13	439	41.8	BG109874	BG109874 602279542	187
14	436	41.5	AA373750	AA373750 zc12903.r1	135
15	432	41.1	W37670	W37670 qc73c11.x	274
16	427	40.7	AA319809	AA319809 qc73c11.x	273
17	425	40.5	AA319809	AA319809 EST363768	271
18	424	40.4	AA319809	AA319809 q223e02.x	270
19	417	39.7	AA319809	AA319809 q223e02.x	265
20	413	39.3	AA319809	AA319809 q223e02.x	259
21	413	39.3	AA319809	AA319809 q223e02.x	256
22	411	39.1	AA319809	AA319809 q223e02.x	256
23	410	38.0	AA319809	AA319809 q223e02.x	256
24	408	38.9	AA319809	AA319809 q223e02.x	256
25	407	38.9	AA319809	AA319809 q223e02.x	256
26	405	38.6	AA319809	AA319809 q223e02.x	256
27	403	38.4	AA319809	AA319809 q223e02.x	256
28	403	38.4	AA319809	AA319809 q223e02.x	256
29	395	37.6	AA319809	AA319809 q223e02.x	256
30	394	37.5	AA319809	AA319809 q223e02.x	256
31	394	37.5	AA319809	AA319809 q223e02.x	256
32	392	37.3	AA319809	AA319809 q223e02.x	256
33	392	37.3	AA319809	AA319809 q223e02.x	256
34	392	37.3	AA319809	AA319809 q223e02.x	256
35	383	36.5	AA319809	AA319809 q223e02.x	256
36	376	35.8	AA319809	AA319809 q223e02.x	256
37	375	35.7	AA319809	AA319809 q223e02.x	256
38	375	35.7	AA319809	AA319809 q223e02.x	256
39	372	35.4	AA319809	AA319809 q223e02.x	256
40	368	35.0	AA319809	AA319809 q223e02.x	256
41	368	35.0	AA319809	AA319809 q223e02.x	256
42	363	34.7	AA319809	AA319809 q223e02.x	256
43	363	34.7	AA319809	AA319809 q223e02.x	256
44	362	34.5	AA319809	AA319809 q223e02.x	256
45	354	33.7	AA319809	AA319809 q223e02.x	256
46	354	33.7	AA319809	AA319809 q223e02.x	256
47	353	33.6	AA319809	AA319809 q223e02.x	256
48	348	33.1	AA319809	AA319809 q223e02.x	256
49	346	33.0	AA319809	AA319809 q223e02.x	256
50	342	32.6	AA319809	AA319809 q223e02.x	256
51	341	32.5	AA319809	AA319809 q223e02.x	256
52	340	32.4	AA319809	AA319809 q223e02.x	256
53	339	32.3	AA319809	AA319809 q223e02.x	256
54	335	31.9	AA319809	AA319809 q223e02.x	256
55	331	31.5	AA319809	AA319809 q223e02.x	256
56	329	31.3	AA319809	AA319809 q223e02.x	256
57	329	31.3	AA319809	AA319809 q223e02.x	256
58	327	31.1	AA319809	AA319809 q223e02.x	256
59	327	31.1	AA319809	AA319809 q223e02.x	256
60	324	30.9	AA319809	AA319809 q223e02.x	256
61	322	30.7	AA319809	AA319809 q223e02.x	256
62	319	30.4	AA319809	AA319809 q223e02.x	256
63	318	30.3	AA319809	AA319809 q223e02.x	256
64	317	30.2	AA319809	AA319809 q223e02.x	256
65	316	30.1	AA319809	AA319809 q223e02.x	256
66	315	30.0	AA319809	AA319809 q223e02.x	256
67	313	29.8	AA319809	AA319809 q223e02.x	256

141	c	141	198	18.7	329	23	A1702447	t294b04.x	214	47	4.5	395	10	AA645356	AA645356	v879q12.r	AA645356	v879q12.r
142	c	198	18.7	196	22		A1589478	tr76f03.x	215	47	4.5	408	3	AA172448	AA172448	ms43f07.r	AA172448	ms43f07.r
143	c	195	18.6	442	159		NA1531	yy04e08..r1	216	47	4.5	410	113	AAW529031	AAW529031	um94e09.y	AAW529031	um94e09.y
144	c	193	18.4	691	195		B6548534	602576527	217	47	4.5	437	141	BE881264	BE881264	y	BE881264	y
145	c	192	18.3	362	189		T61972	y9b6f03..r1	218	47	4.5	445	123	AAW988918	AAW988918	ug12a08.y	AAW988918	ug12a08.y
146	c	190	18.1	324	5		AA298846	EST114447	219	47	4.5	451	168	BF719294	BF719294	mb38f12..	BF719294	mb38f12..
147	c	190	18.1	335	158		H65558	yv67g05..r1	220	47	4.5	460	156	C88814	C88814	Mous	C88814	Mous
148	c	149	18.1	473	190		W52608	z6c9g06..r1	221	47	4.5	468	122	AAW912002	AAW912002	uf40h04.y	AAW912002	uf40h04.y
149	c	189	18.0	310	19		A1349546	qp73610..x	222	47	4.5	469	12	AA839653	AA839653	y	AA839653	y
150	c	189	18.0	342	5		AA298119		223	47	4.5	469	24	A1789021	A1789021	uk55o04.y	A1789021	uk55o04.y
151	c	187	17.8	421	188		R79321	y190c02..s1	224	47	4.5	470	12	AA815658	AA815658	va77h02.r	AA815658	va77h02.r
152	c	185	17.6	185	103		A1880032	ap24b06.x	225	47	4.5	474	12	AA815658	AA815658		AA815658	
153	c	185	17.6	300	5		AA298852	EST114452	226	47	4.5	481	103	AA823064	AA823064		AA823064	
154	c	184	17.5	460	32		AA298852	AV726981	227	47	4.5	481	103	AA823064	AA823064		AA823064	
155	c	183	17.4	593	32		AV717337	AV717337	228	47	4.5	497	17	A1226594	A1226594	uj36e03.y	A1226594	uj36e03.y
156	c	180	17.1	478	188		R77590		229	47	4.5	498	12	AA822963	AA822963		AA822963	
157	c	177	16.9	515	158		H59176	y166a11..r1	230	47	4.5	502	13	AA867880	AA867880	vx22c09.r	AA867880	vx22c09.r
158	c	174	16.6	218	17		A1244079	qj63f02..x	231	47	4.5	502	166	BE303448	BE303448		BE303448	
159	c	173	16.5	674	155		BG573861	602594993	232	47	4.5	503	143	BF011672	BF011672		BF011672	
160	c	172	16.4	957	153		BG426572	602493487	233	47	4.5	505	120	AAW741947	AAW741947		AAW741947	
161	c	169	16.1	423	13		AA328947	o031e03.s	234	47	4.5	505	141	BE853800	BE853800		BE853800	
162	c	166	15.8	427	187		R24623		235	47	4.5	508	114	AAW18610	AAW18610		AAW18610	
163	c	163	15.5	412	12		AA844543	aj21f12.s	236	47	4.5	517	11	AA755546	AA755546		AA755546	
164	c	157	15.0	483	158		H56745	y999e05..r1	237	47	4.5	526	109	A1931504	A1931504	vs77e09.r	AA755546	vs77e09.r
165	c	155	14.8	954	172		BE969051	602266980										











REFERENCE 1 (bases 1 to 681)  
 AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cga@bbs-remail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: CLONTECH Laboratories, Inc.  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Plate: L10M1525 row: 1 column: 18  
 High quality sequence stop: 681.

FEATURES  
 source  
 1. .681  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4597177"  
 /clone\_lib="NIH\_MGC\_61"  
 /tissue\_type="Embryonal carcinoma"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site\_1: SfiI (ggccgctcgcc); Site\_2: SfiI (ggccatcgcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-ATTCTAAGAGCGCCGCGCATG-3' (30)BN-3' (where B = A, C, G or T) Average insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC library."

BASE COUNT 176 a 149 c 157 g 199 t  
 ORIGIN

Query Match 62.1%; Score 652; DB 154; Length 681;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 652; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

242 aggaagatgaacaactgctaccgagcagatgattggtcccttggaagt 301  
 |||||||  
 24 AGGAGTGAACAACCTGGCTCTACCCGACGAGTGTATTTCTTGGAGATG 83  
 |||||||  
 302 cagctccttctgctctgacagtggtgaattccatccctcttattctgagaaa 361  
 |||||||  
 84 CAGTTCCTTTAGTGTGACAGTGTGCAAAATTCATCTCTTATTTCTGAGGAAA 143  
 |||||||  
 362 ctccgtctgctgctgctgctccgagtgagaaagatgatatgtaggaagaa 421  
 |||||||  
 144 CTCCTGTTGTTGGAGTGTGCTGCCAGTGAAGAAAGTGTATGATGATGAGGAA 203  
 |||||||  
 422 actcattgttttaagaaccttcaagctctgagcagctccgtaactgctgtaag 481  
 |||||||  
 204 ACTGAGTGTAAACCTTTCAGTCACTTGGCACTCCGTAATTCCTGTTTCAAG 263  
 |||||||  
 482 aaactctgctcagctaccccccctcaatctctgagtagaacaatgaattgacc 541  
 |||||||  
 264 AAAACCTGTCTTCAGTTCACCTCCCTCAATTCCTGAGTATGAGAACATGATGACC 323  
 |||||||  
 542 tgcctcttcttctgaaactgaaagctacatgatatttcaagctgctgctgata 601  
 |||||||  
 324 TGCCTCTTCTTCTGAACTGCAAGTGTACATGATATTTCAGCTTGTCTCTCATTA 383  
 |||||||  
 602 agcatlaagcaaatcatctcccgatattatcaactggagctggaagctggaatg 661  
 |||||||  
 384 AGCATCTACCAAGATCATCTCCGATATTATATCACTGAGAGCTGGAGAGTTGATG 443  
 |||||||  
 662 aaattgggaagatgataagggaagacttgaacaattcagaagatgcttctaagatccttg 721

|||||  
 Db 444 AAATTGGGAAGCGTTTGGGGAAGACTCTAACAATTCAGAGAACTTCTAAGATCTTGG 503  
 |||||||  
 QY 722 ttgaagctctgcaaaagtttgagatgacatgacatgacatgacatgacatgacatg 781  
 |||||||  
 Db 504 TTGAGCGCTCTGCAAAAGTTTGAGATGACATGATGATGATGATGATGATGATGATG 563  
 |||||||  
 QY 782 ttgaagctctgcaaaagtttgagatgacatgacatgacatgacatgacatgacatg 841  
 |||||||  
 Db 564 TTGAGCTTGTGCTGCTGCAAGTGTATGACACCTCCCTCATTTAGGAGAACAGGACTATTC 623  
 |||||||  
 QY 842 ttgaagcaaaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagca 893  
 |||||||  
 Db 624 TTGAGGCAAAACAGGCAAGAACCCAGCAAGTCTCTTATACCTTGATATTA 675  
 |||||||

RESULT 2  
 BC288867  
 LOCUS BC288867 665 bp mRNA  
 DEFINITION 602383808P1 NIH\_MGC\_93 Homo sapiens CDNA clone IMAGE:4513009 5', mRNA sequence.  
 ACCESSION BC288867  
 VERSION BC288867.1 GI:13044136  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Carnivora; Homnidae; Homo.

REFERENCE 1 (bases 1 to 665)  
 AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cga@bbs-remail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Plate: L10M1039 row: 1 column: 02  
 High quality sequence stop: 665.

FEATURES  
 source  
 1. .665  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4513009"  
 /clone\_lib="NIH\_MGC\_93"  
 /tissue\_type="transitional cell papilloma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: Bladder; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and was constructed by Life Technologies. Note: this is a NIH\_MGC library."

BASE COUNT 185 a 128 c 145 g 207 t  
 ORIGIN

Query Match 56.3%; Score 591; DB 175; Length 665;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 591; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 460 ctccgtaatgcgcctgttcaagaacatctgtctgattcaatccctcaattctctg 519  
 |||||||  
 Db 1 CTCGTAATCCGCGTTCAGAAACCTGTTCTGACGTTCACCTCCCTCAATTTCTCTG 60  
 |||||||  
 QY 520 agtaggaacaatgaagttagctgtcttcttctgaaatgaaatgaaatgaaatgaaatg 579  
 |||||||  
 Db 61 AGTAGGAACAAATGAAGTTGACCTGCTTTCTTGTGAACTGCAAGTGTCAATGATATT 120  
 |||||||  
 QY 580 tcaagctgctgctgctacatcatgacaaagatcatctctcgtattatattca 639



Db 121 TCAAGCTGCTGTCTGTCATTAAGCATCTAGCCAGCATCTCTCTGATTTATTTCA 180  
 Qy 640 ctggaactgcaaggttgatggaattggaagcgtatggaagaaactctgaacaattc 699  
 Db 181 ctggaactgcaaggttgatggaattggaagcgtatggaagaaactctgaacaattc 240  
 Qy 700 agaatgactctgaagatcctgttgagcgtctgcaaaattgcaagatgacatgacat 759  
 Db 241 AGAATGCTTTAAATGATCTGTGTGACGCTTGCAAAAGTTTGCAGATGACATGACAGT 300  
 Qy 760 cttaatggtggaatgagcagtgatgagtgatgacatgacatgacatgacatgacat 819  
 Db 301 CTTATGAGTGGAAATGACAGTGTGATGATGATGATGATGATGATGATGATGATGAT 360  
 Qy 820 attggaagaacaggaactatccttgagcgaacaaagcaagaaacccagcaagtcctat 879  
 Db 361 ATTGGAAGCAAGACATCTCTGAGGCAAAACAAAGCAAGCAAGCAAGTCCTAT 420  
 Qy 880 aacctgcatataatgataatgataatgataatgataatgataatgataatgataatgata 939  
 Db 421 AACCTGCAATATGATATATTTGATATCTGCTGCTTTTCAACATGCTACTTTGGATA 480  
 Qy 940 atgactgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 999  
 Db 481 ATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
 Qy 1000 ggaatgataatgataatgataatgataatgataatgataatgataatgataatgata 1050  
 Db 541 GGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 591

RESULT 3  
 BE897106 947 bp mRNA EST 20-OCT-2000  
 LOCUS 601439663f1 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:3924687 5',  
 DEFINITION mRNA sequence.  
 BE897106  
 VERSION BE897106.1 GI:10362234  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: ATCC/DC/DTP  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L1AM9762 row: n column: 16  
 High quality sequence stop: 656.  
 Location/Qualifiers  
 1. 947  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3924687"  
 /clone\_11b="NIH\_MGC\_72"  
 /tissue\_type="melanotic melanoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: skin; Vector: pCMV-SPORE6; Site: 1; NCI;  
 Site: 2; Salt: Cloned unidirectionally. Primer: Oligo dr.  
 Average insert size 2 kb. Library constructed by Life  
 Technologies."  
 BASE COUNT 274 a 155 c 210 g 307 t 1 others  
 ORIGIN

Query Match 54.0%; Score 567; DB 141; Length 947;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 567; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 484 aactcgtctcagttacaccccccaactctcctgagtaggaacaaagtagtaccct 543  
 Db 1 AACTCTGCTCTGATGCTACCTCCCTCAATCTCTGATGATGATGATGATGATGATGAT 60  
 Qy 544 cctctctctcgaactcgaagtgacatgacatgacatgacatgacatgacatgacatgacat 603  
 Db 61 CTTCTCTCTTCTGACCTCAAGTGTGATGATGATGATGATGATGATGATGATGATGAT 120  
 Qy 604 cactgagcaagatcattcctcctgattatattacatgagcagtgagtgagtgagtgag 663  
 Db 121 CATCTGACCAAGCATCTCTCTGATTTATTTCTGAGCTGCTGCTGCTGCTGCTGCTGCT 180  
 Qy 664 attggaagcgtatgaggaagactctgacaactcagagatgactcttaagatcctgtt 723  
 Db 181 ATTGGAAGGCTTATGGGGAAGACTCTGAACAATTCAGAGATGCTTCTTAAGATCCTGTT 240  
 Qy 724 gacgtctgcaaaatttgagagatgacatgacatgacatgacatgacatgacatgacatgacat 783  
 Db 241 GACGCTGCAAAAGCTTGCAGATGACATGATGATGATGATGATGATGATGATGATGAT 300  
 Qy 784 gagttagtcaactgcaatgacatgacatgacatgacatgacatgacatgacatgacatgacat 843  
 Db 301 GAGTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360  
 Qy 844 gaggcaaacgaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 903  
 Db 361 GAGGCAAAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 420  
 Qy 904 gaatctcgtggtttcaacatgacatgacatgacatgacatgacatgacatgacatgacatgacat 963  
 Db 421 GAATATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480  
 Qy 964 attacacctctacatatttggaatgacatgacatgacatgacatgacatgacatgacatgacat 1023  
 Db 481 ATTATACCTCTTACATATTTTGGACATGATGATGATGATGATGATGATGATGATGAT 540  
 Qy 1024 atgacaacccaagatgacatgacatgacatgacatgacatgacatgacatgacatgacat 1050  
 Db 541 ATGACAAACCAAGATTCGAATGAT 567

RESULT 4  
 BG403007 996 bp mRNA EST 12-MAR-2001  
 LOCUS 602418818f1 NIH\_MGC\_93 Homo sapiens cDNA clone IMAGE:4525781 5',  
 DEFINITION mRNA sequence.  
 BG403007  
 VERSION BG403007.1 GI:13296455  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L1AM10432 row: d column: 06

## High quality sequence stop: 702.

FEATURES  
Location/Qualifiers  
1..996

Source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4525781"  
/clone\_lib="NIH\_MGC\_93"  
/issue\_type="transitional cell papilloma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: bladder; Vector: pCMV-SPORT6; Site:1; Not:  
Site:2; Salt: Cloned unidirectionally; oligo-dt primed.  
Average insert size 1.7 kb. Library enriched for  
full-length clones and constructed by lile technologies.  
Note: this is a NIH\_MGC library."

BASE COUNT 279 a 186 c 235 g 296 t  
ORIGIN

Query Match 50.4%; Score 529; DB 153; Length 996;  
Best Local Similarity 100.0%; Pred. No. 1,7e-281;  
Matches 529; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 353 ctggaggaactcctgtgttttgcagttggtccagtgaggaagagtgatagtag 412  
DB 8 CTGGAGGAACCTCTGTGTTGGCAGTTGGCTCCAGAGGAAGAGTGTATAGTAG 67  
QY 413 ggaaggaactcagtgatttgaagaccttcacacctgcccagctccgaatggcc 472  
DB 68 GGAAGGAACCTCAGTGTTCAGACCTTTCACTTCGCGCACTCCGTAATGCC 127  
QY 473 ttttcaagaagaactcgttctcagttcactcccccataattcctgagtagaacaatg 532  
DB 128 TGTTCAGAGAAACCTGTCTCTCAGTTCACTCCCTCAATTCCTGAGTAGAACAAATG 187  
QY 533 aagtgacctgctcttcttctcgaactgaagtgctcaagatattcaagcttgctgt 592  
DB 188 AAGTTGACCTGCTCTTCTTCTGCACTGCAAGTGCTCAATGATTTTCACTTGCTGT 247  
QY 593 ctggtcaatagatcctcgaagaagatattcctcgtatattatcactgagagctgag 652  
DB 248 CTGCTCAATAGCACTCTGCAAGAGATATTCCTGATTTATATTCACGAGCGAG 307  
QY 653 gtttgatgaacttggaagcgtataggggaagactcgaacaattcagaagatgcttcta 712  
DB 308 GTTTGATGAATGGAAGCGTTATGGGGAAGACTCTGAACAAATTCACAGATGCTTCTTA 367  
QY 713 agatccttgaagctctcgaagaagtttgcagatgacatgacatgctttagtggga 772  
DB 368 AGATCCTTGTGACCTCTGCAAAAGTTTGCAGATGACATGACATCTTTATGCTGGGA 427  
QY 773 atgcagtgtagagttagctcgtcgaatcatttgaacacctccatcattgaagaaca 832  
DB 428 ATGAGAGGTAGAGTTAGTACTGTCAGTCAATTTGACACCTCCCTCATTTAGAGAACAA 487  
QY 833 ggaactccttgaagcaaaagaagcaagaacccagcaagtcctcctataa 881  
DB 488 GGACTATCCTTAGGCAAAACAGCAAGCAACCCAGCAAGTCCCTATTA 536

RESULT 5  
LOCUS BF060652 516 bp mRNA EST 16-OCT-2000  
DEFINITION 7.59a04.x1 Soares NSF\_F8\_9W\_OT\_PA\_P-S1 Homo sapiens cDNA clone  
IMAGE:3390702 3 similar to TR:Q9VHG4 Q9VHG4 G88444 PROTEIN. ;  
ACCESSION BF060652  
VERSION BF060652.1 GI:10819562  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 516)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
CONTACT Robert Strausberg, Ph.D.  
COMMENT Email: cgaaps-r@mail.nih.gov  
This clone is available royalty-free through LNL: contact the  
IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
Seq primer: -400p from gibco  
High quality sequence stop: 469.

FEATURES  
source  
1..516  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3390702"  
/clone\_lib="Soares NSF\_F8\_9W\_OT\_PA\_P-S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: p773D-Pac (Pharmacia) with  
a modified polyLinker; Site:1; Not I; Site:2; Eco RI;  
Equal amounts of plasmid DNA from five normalized  
libraries were mixed, and ss circles were made in vitro.  
Following HAP purification, this DNA was used as tracer in  
PCR-amplified cDNAs from pools of 5,000 clones made from  
the same 5 libraries. The pools consisted of the following  
libraries and clones: Soares NBHSF pool 1:  
309384-310919, 323208-325895 Soares NB2HP pool 1:  
145032-147335, 147720-148103, 148872-149255, 15002 -  
150407, 151176-152327 Soares NB2HP-9W pool 1:  
758280-760583, 772104-774407 Soares NBHPA pool 1:  
304776-306311, 320136-322823, 326280-326663 Soares NBHOT  
pool 1: 723720-726407, 739080-740999 Subtraction by Bento  
Soares and M. Fatima Bonaldo."

BASE COUNT 118 a 119 c 125 g 154 t  
ORIGIN

Query Match 49.1%; Score 516; DB 143; Length 516;  
Best Local Similarity 100.0%; Pred. No. 2.6e-274;  
Matches 516; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 ggaagcagattagatataaataacacagagctgtgttctccgaatggaatgg 105  
DB 1 GGAAGCAGATTAGATATAAATACACAGAGCTGTGTGTTCCGAAATGG 60  
QY 106 cctatacagagagcgatccacagacgtgagctgcatgtcagtgagcttctgtaaa 165  
DB 61 CCTATACGAGAGACGGATCCAGACGCTGATTCATTCATGGGCTCTCTGTA 120  
QY 166 gaagaccttcttgccagagctcgaatggtaacctgtttcaatgctccgggctaac 225  
DB 121 GAAGACCTTTCTTGCCAGACACTCGAGTGGGTAACTGTTTCATCTCGGGCTACC 180  
QY 226 gtaatgtagtgaagggagtgaaactggcttaccaccacagagcagtgatattcg 285  
DB 181 GTCATGCTGATGTTGAAGGAGTGAACAACTGCTCTACCCCGACGAGTCAATTGG 240  
QY 286 tacccttggaagatgcaagcttctttagctttagcagtgatggaatcattcactcc 345  
DB 241 TACCTTTGGAATGCAAGTCTTTTATGCTTGACAGTGTTCAGAAATTCATTCACTCC 300  
QY 346 ttatttctggaagaaactcctgtgttttgcagttgctccagtgaggaagagtgat 405  
DB 301 TTATTCTTGAGGAACCTCTGTGTTTGCAGTTGGCTCCAGTGAAGAAAGTGTAT 360  
QY 406 atgtagggaaggaactcagtgatttaaaccttcagtaaccttggcgaagcccgat 465  
DB 361 ATGTTAGGGAAGCAACTCAGTGTTTGAAGCTTTCAGTCACTTCGCGACACTCGT 420  
QY 466 aatgcgctgttcaagaagaactctgttctcagttcactccctcaattctctgtagag 525  
DB 421 AATGCGCTGTTCAAGAAACCTGTCTCAGTTCACTCCCTCAATTCCTGAGTAGG 480

QY 526 aacatgaatgacactgctcttcttctgaactg 561  
 DB 481 AACAAATGAAGTGAAGTCTCTCTCTCTCTGAACCTG 516  
 RESULT 6  
 BG501143 852 bp mRNA EST 27-MAR-2001  
 LOCUS 602546382FI NIH\_MGC\_60 Homo sapiens cDNA clone IMAGE:4668621 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG501143  
 VERSION BG501143.1 GI:13462660  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 852)  
 NIH-MGC http://mgs.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL Contact: Robert Strausberg, Ph.D.  
 COMMENT Email: cgapds-remail.nih.gov  
 Tissue Procurement: DCTD/DTF  
 cDNA Library Preparation: CLONETECH Laboratories, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LICM1479 row: c column: 22  
 High quality sequence stop: 781.  
 Location/Qualifiers  
 source 1..852  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="4668621"  
 /clone\_lib="NIH\_MGC\_60"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /tissue\_type="adrenocarcinoma"  
 /note="Organ: prostate; Vector: pDNR-LIB (Clontech);  
 Site\_1: SfiI (ggcgctcgcc); Site\_2: SfiI (ggcgctcgcc)  
 ; Double-stranded cDNA was prepared from cell line RNA.  
 5' and 3' adaptors were used in cloning as follows: 5'  
 adaptor sequence: 5'-CACGGCCATTATGCGC-3' and 3' adaptor  
 sequence: 5'-ATTCTAGAGCCGCGCGCGCGCGCATG-3' (30)BN-3'  
 (where B = A, C, G, or T). Average  
 insert size 1.5 kb (range 0.9-4.0 kb). 14/15 clones for  
 contained inserts by PCR. This library was enriched for  
 full-length clones and was constructed by Clontech  
 Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC  
 library."  
 BASE COUNT 256 a 131 c 183 g 282 t  
 ORIGIN  
 Query Match 47.5%; Score 499; DB 154; Length 852;  
 Best local Similarity 100.0%; Pred. No. 7e-265;  
 Matches 499; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 181 GCAAAAGTTGGAGATGACATGTACATGTTTATGTTGGAGATGACATGTTAGTTAGT 240  
 QY 792 cactgcaagtcatttgacacccctccatlaaggaagaagacatctcttgaggcaaa 851  
 DB 241 CACTGTCAATGATTTGACACCTCCATCATTAGGAAGACAGACATCTCTTGGAGCAAA 300  
 QY 852 acaagcgaagaccgagcgaagtccttaacatcttgacataagtaatttgatattc 911  
 DB 301 ACACCGAAGAACCCACAGATCCCTATACCTGTCATATAGTATATTTTGAATATTC 360  
 QY 912 cgtggtttcaacatggtacttggaataagatgccttgcccttgctgctgattatcc 971  
 DB 361 CGTGTTTCAACATGTTGACTTTGATATGATGATGATGATGATGATGATGATGATGAT 420  
 QY 972 ctctacacatatttggaacatgacatccttgatagatgacatcattatagatgacaa 1031  
 DB 421 CTCTTACATATTTGGAACTGATCTCGATATGATGATGATGATGATGATGATGATGAT 480  
 QY 1032 ccagaagatcgaattgat 1050  
 DB 481 CCAGAAGATTCGAATGAT 499  
 RESULT 7  
 BF196461 480 bp mRNA EST 03-NOV-2000  
 LOCUS BF196461  
 DEFINITION hr86e07.x1 NCI-CGAP Kid1 Homo sapiens cDNA clone IMAGE:3135396 3',  
 similar to TR:Q9VHG4 Q9VHG4 C68444 PROTEIN.; mRNA sequence.  
 ACCESSION BF196461  
 VERSION BF196461.1 GI:11084432  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 480)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP).  
 TITLE Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapds-remail.nih.gov  
 Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Sequencing by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL, send email to:  
 info@image.llnl.gov  
 High quality sequence stop: 435.  
 Location/Qualifiers  
 source 1..480  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="3135396"  
 /clone\_lib="NCI-CGAP\_Kid1"  
 /lab\_host="DH10B"  
 /note="Organ: Kidney; Vector: pRT73D-Pac (Pharmacia) with  
 a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
 Plasmid DNA from the normalized library NCI-CGAP Kid1 was  
 prepared, and ss circles were made in vitro. Following HAP  
 purification, this DNA was used as tracer in a subtractive  
 hybridization reaction. The driver was PCR-amplified cDNAs  
 from a pool of 5,000 clones made from the same library  
 (clonoids 1322376-1323911, 1456007-1456775, and  
 1500552-1502855). Subtraction by Bento Soares and M.  
 Fatima Bonalde."  
 BASE COUNT 109 a 111 c 119 g 141 t  
 ORIGIN

Query Match 45.7%; Score 480; DB 145; Length 480;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-254;  
 Matches 480; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 gggagagattatataataaacaccaggtcgtgtgttcttcgaatgaaatg 105  
 Db 1 GGGAGAGATTATATAATAAACACCAGGTCGTGTGTTCGAAATGGAAATGG 60  
 QY 106 cctataccagagagagatccagagctgctgcatctgtccatggtctctgtaa 165  
 Db 61 CCTATACCAAGAGAGAGATCCAGACCTGGCTGCATGTCTCATGGCTCTCTGAA 120  
 QY 166 gaagacccctcttgccagagctgcagctggaactcttcaatctgctctggttacc 225  
 Db 121 GAAGACCTTCTTGCCAGAGATCGCACTGGGTAACTCTTTCATCTCTCGGGCTACC 180  
 QY 226 gtcacgtgtgtagtgaagggagtgacaacatgctctacccacagagctgcatctg 285  
 Db 181 GTCAGGTGATGATGTAAGGAGATGACAACTGGCTCTACCCAGGACGTGATTCG 240  
 QY 286 tacctcttgagagatcagctccttcttgaactgctggaatgttgaatcttacc 345  
 Db 241 TACCTTTGGAGATGCAATTCCTTTAGCTTGACAGTGTGCAAAATTCATTCATCC 300  
 QY 346 ttattcttgagagaaactcctcgtgtgttgcagtgctcctcagtgagaaagagtgat 405  
 Db 301 TTATTTTCTGAGGAAACCTGCTGTGTTTCCAGTTGCTCCAGTGAAGAGAGTAT 360  
 QY 406 atggtagagagagcaactcagtggttgagagaccttgaacacttgagagagagtgat 465  
 Db 361 ATGTTAGGAGAGCAAACTCAGTGTGTAAGACCTTTGACGACCTTTGGCCAGCTCGT 420  
 QY 466 aatgcctgttctcaagaaactcgtctcagttcaactcctccatcattcttgtag 525  
 Db 421 AATGCGCTGTTCAAGAAACTGCTGTTCTCAGTTCACCTCCCTCAATTCCTGAGTGG 480

RESULT 8  
 AM166105 478 bp mRNA EST 12-NOV-1999  
 LOCUS x145609.x1 NCI-CGAP\_Brn50 Homo sapiens cDNA clone IMAGE:2621008 3'  
 DEFINITION mRNA sequence.  
 ACCESSION AM166105  
 VERSION AM166105.1 GI:6397630  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI/NCI/NCBI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute / National Institute of Neurological  
 Disorders and Stroke, Brain Tumor Genome Anatomy Project  
 (CGAP/RTGAP), Tumor Gene Index  
 Unpublished (1998)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-remail.nih.gov  
 Tissue Procurement: Burt Feuerstein, M.D., Mark Israel, M.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 found through the I.M.A.G.E. Consortium/LNW at:  
 www.bio.linn.gov/dbp/image/image.html  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 422.  
 Location/Qualifiers  
 1.478  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2621008"  
 /clone\_id="NCI-CGAP\_Brn50"  
 /tissue\_type="medulloblastoma"

/lab.host="DH10B (phage resistant)"  
 /note="Organ: brain; Vector: p773D-Pac (Pharmacia) with a  
 modified polylinker; 1st strand cDNA was prepared from  
 medulloblastoma tumor tissue, and was then primed with a  
 Not I - oligo(dT) primer. Double-stranded cDNA was ligated  
 to Eco RI adaptors (Pharmacia), digested with Not I and  
 cloned into the Not I and Eco RI sites of the modified  
 p773 vector. This library is normalized. Library  
 constructed by Bento Soares and M. Fatima Bonaldo."  
 BASE COUNT 109 a 111 c 117 g 141 t  
 ORIGIN

Query Match 45.5%; Score 478; DB 112; Length 478;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-253;  
 Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 gggagagattatataataaacaccaggtcgtgtgttcttcgaatgaaatg 105  
 Db 1 GGGAGAGATTATATAATAAACACCAGGTCGTGTGTTCGAAATGGAAATGG 60  
 QY 106 cctataccagagagagatccagagctgctgcatctgtccatggtctctgtaa 165  
 Db 61 CCTATACCAAGAGAGAGATCCAGACCTGGCTGCATGTCTCATGGCTCTCTGAA 120  
 QY 166 gaagacccctcttgccagagctgcagctggaactcttcaatctgctctggttacc 225  
 Db 121 GAAGACCTTCTTGCCAGAGATCGCACTGGGTAACTCTTTCATCTCTCGGGCTACC 180  
 QY 226 gtcacgtgtgtagtgaagggagtgacaacatgctctacccacagagctgcatctg 285  
 Db 181 GTCAGGTGATGATGTAAGGAGATGACAACTGGCTCTACCCAGGACGTGATTCG 240  
 QY 286 tacctcttgagagaaactcctcgtgtgttgcagtgctcctcagtgagaaagagtgat 345  
 Db 241 TACCTTTGGAGATGCAATTCCTTTAGCTTGACAGTGTGCAAAATTCATTCATCC 300  
 QY 346 ttattcttgagagaaactcctcgtgtgttgcagtgctcctcagtgagaaagagtgat 405  
 Db 301 TTATTTTCTGAGGAAACCTGCTGTGTTTCCAGTTGCTCCAGTGAAGAGTAT 360  
 QY 406 atggtagagagagcaactcagtggttgagagaccttgaacacttgagagagagtgat 465  
 Db 361 ATGTTAGGAGAGCAAACTCAGTGTGTAAGACCTTTGACGACCTTTGGCCAGCTCGT 420  
 QY 466 aatgcctgttctcaagaaactcgtctcagttcaactcctccatcattcttgtag 523  
 Db 421 AATGCGCTGTTCAAGAAACTGCTGTTCTCAGTTCACCTCCCTCAATTCCTGAGTGA 478

RESULT 9  
 BF732377 452 bp mRNA EST 30-MAR-2001  
 LOCUS nae11603.x1 NCI-CGAP\_OY18 Homo sapiens cDNA clone IMAGE:3434788 3'  
 DEFINITION similar to TR:05YHG4 09YHG4 CG8444 PROTEIN.; mRNA sequence.  
 ACCESSION BF732377  
 VERSION BF732377.1 GI:12057535  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-remail.nih.gov  
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
 R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: M. Bento  
 Soares, Ph.D. CDNA Library Arrayed by: Christa Prange, The  
 I.M.A.G.E. Consortium DNA Sequencing by: Washington University

## FEATURES

source  
 1.478  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2621008"  
 /clone\_id="NCI-CGAP\_Brn50"  
 /tissue\_type="medulloblastoma"

Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL, send email to:  
 info@image.llnl.gov  
 Seq primer: -400P from Gibco  
 High quality sequence stop: 437.  
 Location/Qualifiers

## FEATURES

source

1. 452  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3434788"  
 /clone\_lib="NCI CGAP OV18"  
 /tissue\_type="fibrothecoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: ovary; Vector: p773D-Pac (Pharmacia) with a  
 modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5'  
 TGTACCAATCTGAGTGGAGCGCGCCGCGACATTTTCTTTTCTTTTCTT  
 3'];  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not  
 I and Eco RI sites of the modified p773 vector. Library  
 went through one round of normalization, and was  
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 104 a 101 c 115 g 132 t  
 ORIGIN

Query Match 43.0% Score 452; DB 168; Length 452;  
 Best Local Similarity 100.0% Pred. No. 7.2e-239;

Matches 452; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

46 gggaacagattagatataaataccagagctgtgttttcggaatggaattgg 105  
 1 GGGAAACAGATTAGTATATTAATCAACAGGCTCTGTTTCCGAATGGAAATTGG 60  
 106 cctaccagagagagagatccagagctgtgttcacatggtcctctctgtgaa 165  
 61 CCTATACCAAGAGAGGATCCAGAGCTGCTGATGCTCATGGGCTCTCTGTGAAA 120  
 166 gaagacattcttggcagagctgcagtggttaacctgttcatgctctgtgctac 225  
 121 GAAAGACCTTTCTGGCAGAGCTCGAGTGGGTACCTGTTTCATCTCTCGGCTAC 180  
 226 gtcatgtgtatgttgaaggagtgaaacaactgtgctctaccccaaggcaatgtc 285  
 181 GTCATGTGTATGCTGAAGAGATGAACTGCTTACCCCAAGGCAATGTCTTTCG 240  
 286 taccttggagaatgcaatcttcttgaatgttgaatgttgaatgttgaatgtt 345  
 241 TACCTTTGGAGAAATGCACTTCTTTAGCTTTCAGACAGTGGCAAAATTCATTC 300  
 346 ttttttcttgaagaaactctgttcttcttgaatgttgaatgttgaatgttgaat 405  
 301 TTTATTTCTGAGAAATCTCTGTTTTCAGTGTGCTCCAGTGAAGAAAGTGTAT 360  
 406 atgttgaaggaagcaactcagtggttgaagaccttgaatgttgaatgttgaat 465  
 361 ATGTGAGGGAAGCAACCTCAGTGTGAAACCTTTCAGTACCTTGGCGCAGCTCG 420  
 466 aatcgctgttcaagaagaactgttctcag 497  
 421 AATCGCTGTTCAGAAACCTGTCTCTCAG 452

RESULT 10  
 AM024459 501 bp mRNA EST 09-MAR-2000  
 LOCUS AM024459 wu/6f06.x1 NCI CGAP\_Kid3 Homo sapiens cDNA clone IMAGE:2525987 3',  
 DEFINITION mRNA sequence.  
 ACCESSION AM024459  
 VERSION AM024459.1 GI:5877989  
 KEYWORDS EST.  
 SOURCE human.

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgsb@remail.nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www.blo.llnl.gov/bdrp/image/image.html  
 Insert Length: 644 Std Error: 0.00  
 Seq primer: -400P from Gibco  
 High quality sequence stop: 425.  
 Location/Qualifiers

## FEATURES

source

1. 501  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2525987"  
 /clone\_lib="NCI CGAP\_Kid3"  
 /lab\_host="DH10B"  
 /note="Organ: kidney; Vector: p773D-Pac (Pharmacia) with  
 a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer,  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not  
 I and Eco RI sites of the modified p773 vector. mRNA  
 source: 2. pooled kidneys. Library went through one round  
 of normalization. Library constructed by Bento Soares and  
 M. Fatima Bonaldo."

BASE COUNT 115 a 116 c 123 g 147 t  
 ORIGIN

Query Match 42.9% Score 450; DB 110; Length 501;  
 Best Local Similarity 99.8% Pred. No. 9.3e-238;

Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

46 gggaacagattagatataaataccagagctgtgttttcggaatggaattgg 105  
 1 GGGAAACAGATTAGTATATTAATCAACAGGCTCTGTTTCCGAATGGAAATTGG 60  
 106 cctaccagagagagagatccagagctgtgttcacatggtcctctctgtgaa 165  
 61 CCTATACCAAGAGAGGATCCAGAGCTGCTGATGCTCATGGGCTCTCTGTGAAA 120  
 166 gaagacattcttggcagagctgcagtggttaacctgttcatgctctgtgctac 225  
 121 GAAAGACCTTTCTGGCAGAGCTCGAGTGGGTACCTGTTTCATCTCTCGGCTAC 180  
 226 gtcatgtgtatgttgaaggagtgaaacaactgtgctctaccccaaggcaatgtc 285  
 181 GTCATGTGTATGCTGAAGAGATGAACTGCTTACCCCAAGGCAATGTCTTTCG 240  
 286 taccttggagaatgcaatcttcttgaatgttgaatgttgaatgttgaatgtt 345  
 241 TACCTTTGGAGAAATGCACTTCTTTAGCTTTCAGACAGTGGCAAAATTCATTC 300  
 346 ttttttcttgaagaaactctgttcttcttgaatgttgaatgttgaatgttgaat 405  
 301 TTTATTTCTGAGAAATCTCTGTTTTCAGTGTGCTCCAGTGAAGAAAGTGTAT 360  
 406 atgttgaaggaagcaactcagtggttgaagaccttgaatgttgaatgttgaat 465  
 361 ATGTGAGGGAAGCAACCTCAGTGTGAAACCTTTCAGTACCTTGGCGCAGCTCG 420

QY 466 aatcgccgttcaagaacacatctgtctcaatcccccctcaaatctctcagtagg 525  
|||||  
Db 421 AATGCCCTGTTTCAAGATACCTGTCTCACTCACTCCCTCAATCTCTGATAGG 480  
QY 526 aacatgaagttagctgtc 546  
|||||  
Db 481 AACATGAAGTTGACCTGCTC 501

## RESULT 11

AM003884 448 bp mRNA EST 08-SEP-1999  
LOCUS w61f04.x1 NCL\_CGAP\_Brn23 Homo sapiens cDNA clone IMAGE:2501695 3',  
DEFINITION mRNA sequence.  
ACCESSION AM003884  
VERSION AM003884.1 GI:5850800  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 448)  
AUTHORS NCI/NINDS-CGAP NCL:www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute / National Institute of Neurological  
Disorders and Stroke, Brain Tumor Genome Anatomy Project  
(CGAP/BRGAP), Tumor Gene Index  
Unpublished (1998)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: cgapbs-rt@mail.nih.gov  
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,  
Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/bdrrp/image/image.html  
Seq primer: -40up from Gibco  
High quality sequence stop: 445.  
Location/Qualifiers  
1. 448  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="IMAGE:2501695"  
/clone\_lib="NCL\_CGAP\_Brn23"  
/tissue\_type="Glioblastoma (pooled)"  
/lab\_host="DH10B"  
/note="Organ: Brain; Vector: p7T3D-Pac (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTCACCAATCTGAAGTGGAGCGCGCATATCTTTTCTTTTCTTTTCTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia). digested with Not I and cloned into  
the Not I and Eco RI sites of the modified p7T3 vector.  
Library is normalized, and was constructed by Bento  
Soares and M. Fatima Bonaldo."

## FEATURES

BASE COUNT 99 a 98 c 119 g 132 t  
ORIGIN  
Query Match 42.7%; Score 448; DB 110; Length 448;  
Best Local Similarity 100.0%; Pred. No. 1.2e-236;  
Matches 448; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 34 gggggtgtttggggaagagtttagatataaataccacaggtctgtgttttcga 93  
|||||  
Db 1 GCGGGGTGTTTGGGGAACGAGTTTACTATATTAATAACACAGAGGTGTTGTTCCGA 60  
QY 94 aatggaatgtcctataccagagagagatcccaagctgtgtgttcacatggc 153  
|||||  
Db 61 AATGGAATGTGCTATACAGAGAGCGGATCCCAAGCTGCTGCTGCTGCTGCTGCTG 120

QY 154 ttctctgtgaagaagaccttcttggccagagctgcagtgagglaaacgtttacatg 213  
|||||  
Db 121 TTCTCTGTGAAGAAGAACCTTCTTGGCCAGAGACTCGAGTGGGTAACTGTTTATGCT 180  
QY 214 cctcgagctaccgctcaatgtagtgaagagagtgaaacaaatgcgtctacccccagc 273  
|||||  
Db 181 CCTGGGCTACCGCTCATGTGTATGTGTGAAGAGAGTGAACAACTGCTTACCCCAAGC 240  
QY 274 agtgcatttcgtaaccttggagaatgcagttccttttagctctgacagtgtaaat 333  
|||||  
Db 241 AATGTCATTTGACCTTTGAGAAATGACAGTCTCTTTAGCTTTCAGATGTGCAAT 300  
QY 334 tccatcaactccttatttctcgaagaacccctgtgtgttgcagttgtgtccagtag 393  
|||||  
Db 301 TCATATCAGTCTTATTTTCTGAGGAACCTCCTGTTTGTGCACTTGCTCCAGTAG 360  
QY 394 gaaagatgtatatagttggaggaagcaaacatgcagtttgaagaccttcaaccttg 453  
|||||  
Db 361 GAAAGAGTGTATGTGTAGGAGGAGCAAACTCAGTGTGTAAGACCTTTCAGTCACTTG 420  
QY 454 cgcagctccgtaacgcctgtttcaag 481  
|||||  
Db 421 CGCCAGCTCGTATATCGCTGTTCAAG 448

## RESULT 12

AM958900 624 bp mRNA EST 01-JUN-2000  
LOCUS EST370970 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.  
DEFINITION  
ACCESSION AM958900  
VERSION AM958900.1 GI:8148584  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 624)  
AUTHORS Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspar, R., Gay, C., Holt  
I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and  
Quackenbush, J.  
Assessment of gene expression patterns in a model of colon tumor  
metastasis using a 19,200 element cDNA microarray  
Unpublished (2000)  
CONTACT: John Quackenbush  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 3528  
Fax: 301 838 0208  
Email: johnq@tigr.org  
Plate: 127  
Seq primer: Reverse.  
Location/Qualifiers  
1. 624  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="MAGE resequences, MAGE"  
/note="Vector: pBluescriptSKm"

BASE COUNT 162 a 135 c 134 g 193 t  
ORIGIN

Query Match 41.9%; Score 440; DB 122; Length 624;  
Best Local Similarity 99.6%; Pred. No. 3.3e-232;  
Matches 540; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 278 tcaatctacccttgaggaatgcagttccttttagtctgaagagtggttgaattcca 337  
|||||  
Db 1 TCATTTCTGATCCCTTTTGGAATGAGATGCTTTTATGCTGTGACAGTGTGCAAAATTCGA 60  
QY 338 ttaactccttattcttgagaagaacctcgtgtgttgcagttgtgtcccaagtagaaga 397  
|||||  
Db 61 TTCACTCTTATTTTCTGAGGAACCTCCTGTTGTTTGGAGTGTGCTGCCAGTAGAGAAA 120









Db	1	GGGAAGCACTTTAGTATATTAATAATCCAGGAGCTCTGTTTCCGAATGGAAATTTG	60
OY	106	ccataccagagagacgagatcccaagacgtgctcatgtccatggctctctctgfgaaa	16
Db	61	CGATACGAGAGACCGGATCCGACAGCGGCTCATGTTCATGAGCTTCTCTGTGAAA	12
OY	166	gaagaccttcttggccaagactcgcagtggtglaacctgttcaatcgtccctcaggtacc	22
Db	121	GAAGACCTTCTTTGGCCAGAGACTCGAGTGGGTAACTGTTTCATGTCCTCGGGCTACC	18
OY	226	gtaatgtagtbfagaagggagatgaacaaactgtcttaacccagcgaatgtcaattcg	28
Db	181	GCTATGATATGTGTGAAGGAGATGAACAACCTGGCTCTTACCCCGAGCGATGTCATTGG	24
OY	286	tacccttggagaaatgcagttccctttagctcttgacagatgttgcaaatccatcaatcc	34
Db	241	TACCTTTGGAGAAATGCAGTTCCTTTAGTCTTACAGTGTTCGAATTCATTCACTCC	30
OY	346	ttaatttctcggagaaactcctctgtgtcttcttgaaagtgtgctcccaatgagaagaagtat	40
Db	301	TATTTTCTCGAGAACTCCTGTTGTTTGGACGTGGCTCCACGAGAAAGAGTAT	36
OY	406	atgtgtaggaaggaactcagtggtcttgaacacttcaagttaaccttgcgcagctccgt	46
Db	361	ATGTGAGGAGAGCAACCTCAGTGTGTTGAAGACCTTTCAGTCACTTCGCCGACGCTCGT	42
OY	466	aatcg 470	
Db	421	AATCG 425	

RESULT	18
AM951698	
LOCUS	AM951698 424 bp mRNA
DEFINITION	EST363766 MAGE rescueses, MACB Homo sapiens CDNA, mRNA sequence
ACCESSION	AM951698
VERSION	AM951698.1 GI:8141377
KEYWORDS	EST.
SOURCE	human.

ORGANISM	REFERENCE	AUTHORS
Homo sapiens		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Mammalia; Euteria; Primates; Carnivora; Hominoidea; Homo.		
1 (basses 1 to 424)		
Heyde, P., Qi, R., Aernathy, K., Bharad, S., Gaspard, R., Gav, C., Holt,		

TITLE	I.E., Saeed A.I., Sharov, V., Lee, N.H., Yeatman, T.J., and Quackenbush, J.
JOURNAL	Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray
COMMENT	Unpublished (2000) Contact: John Quackenbush

The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 3528  
Fax: 301 838 0208  
Email: [johnget@igr.org](mailto:johnget@igr.org)  
Plate: 30  
Seq primer: Reverse.  
Location/Qualifiers

```

SOURCE      1. . 424
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone_lib="MAGE resequences, MAGB"
             /note="Vector: pBluescriptSkm"
BASE COUNT  109 a      95 c      92 g      128 t
ORIGIN

```

Query Match	40.48%	Score 424	DB 122	Length 424	
Best Local Similarity	100.0%	Pred. No. 2	3e-223		
Matches 424	Conservative 0	Mismatches 0	Indels 0	Gaps 0	
418	gcaaacctcagtttgaagaccttaagtaacacttcggccagctccgtaacgcgtgttt 477				

Db	1	GCAAACTAGTGTGGAAGACCTTTCAGTACCTTGCCAGACTCCGTAATGACCTGTTT	60
OY	478	caagaaacatctgtctcagattcactccccctaattctcgtatgagaacaagaatc	537
Db	61	CAGGAACCTCTGTTCAGATTCACCTCCCTCAATTCCTGAGTAGAACAATACMTT	120
OY	538	gaacctccttcttcttctgaactgcaagtgatcacaatattcaagctgtcgtcgt	597
Db	121	GACCTGCTCTTCTTCTTCACTGAACGTGACGTCTCATGATTAATTCACCTGCTGCTGCT	180
OY	598	cataagacatcagacgaagatcatctccatattatattcaatgagatctgcagattg	657
Db	181	CATAAGCATCTAGSCAAGATATCTCTCGATTATATATACGAGCGCTGCACAGTTTG	240
OY	658	gataaatttggaagcgtattatgggaagactctgaaacattcagaatctgcttaagtc	717
Db	241	GATGAATATGGGAAGCGTTATGGGGAAGACTCGAACAATTCAAGATGCTTCTTAAGTC	300
OY	718	ctcttgtagcgtctcgaaaagtctgcagatgacatgacatcagctcttattgvtggaaatgca	777
Db	301	CTCTGTGACGCTCTGGAAAAGTTTGCAGATGACATGTCACACTCTTATATGTTGGGAATGCA	360
OY	778	gtcgtatagattacatctcaatcaatcttgaaacctccctcatatggagaacaaggact	837
Db	361	GTCGTAGAGTTAGTACACTCATCACTCAATTTGAACTCTCCCATNTAGGAAGCAAGACT	420
OY	838	atcc 841	
Db	421	ATCC 424	

RESULT	19						
LOCUS	AI366997						
DEFINITION	AI366997	417 bp	mrna	EST	16-FEB-1999		
ACCSSION	q236e2.x1	NCI_CGAP_CLL1	Homo sapiens	CDNA clone	IMAGE:202738	3'	
VERSION	AI366997						
KEYWORDS	AI366997.1	GI:4136742					
SOURCE	human.						
ORGANISM	Homo sapiens						

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukariyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 417)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.

Email: [cgapds-remail.nih.gov](mailto:cgapds-remail.nih.gov)  
Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever, M.D., Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www.bio.llnl.gov/bdrrp/image/image.html](http://www.bio.llnl.gov/bdrrp/image/image.html)  
Insert Length: 1913 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 350.  
Location/Qualifiers  
1.417

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
clone_image="2027738"
clone_1b="NCI CGAP CLL1"
tissue_type="B-cell, chronic lymphocytic leukemia"
lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand
cDNA was primed with a Not I - oligo(dT) primer [5'

```



JOURNAL  
COMMENT

Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Clone distribution: NCI-CCAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
www.bio.lnl.gov/bdrip/image/image.html  
Seq primer: -40m13 fwd. RT from Amersham  
High quality sequence stop: 433.

## FEATURES

## Source

Location/Qualifiers  
1. 458  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1605403"  
/clone\_lib="NCI-CCAP Kid5"  
/tissue\_type="2 pooled tumors (clear cell type)"  
/lab\_host="DH10B"  
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
AAGTGAAGAATTCGGCGCCGCAATATTTTTTTTTTTT 3']  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT  
ORIGIN

104 a 104 c 115 g 135 t

Query Match 39.3%; Score 413; DB 14; Length 458;  
Best Local Similarity 100.0%; Pred. No. 2.9e-217;  
Matches 413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 46 gggaagaggtttagtataataataaccagggtctgttcttcggaatgaaattgg 105  
DB 1 gggagacgagtttagtattttaaataccagggctgtgtttccgaatgcaaatggc 60  
OY 106 cctataccagagacgagatccagagcgtgctgcatgtccatgggtctctgtgaaa 165  
DB 61 cctatccagagacgagatccagagcgtgctgcatgtccatgggtctctgtgaaa 120  
OY 166 gaagaccttcttgcagagactcgcagtggttaacctgttcatcgtctcgggtacc 225  
DB 121 gaagaccttcttgcagagactcgcagtggttaacctgttcatcgtctcgggtacc 180  
OY 226 gtaatgtagtgaagagagtgaaacaactgagcttaccacccagagcagtgatcttg 285  
DB 181 gtcagtgtagtgaagagagtgaaacaactgagcttaccacccagagcagtgatcttg 240  
OY 286 tacccttggagaaagcagcttctttagtcttgcagtggttgcgaatctcaactacc 345  
DB 241 tacccttggagaaagcagcttctttagtcttgcagtggttgcgaatctcaactacc 300  
OY 346 ttatttctgaagaaactcctgtgttctgcagtggtcctccagtgagaaagagtgat 405  
DB 301 ttatttctgaagaaactcctgtgttctgcagtggtcctccagtgagaaagagtgat 360  
OY 406 atgtgtggaagaaagcagtgatgttgaagaaccttcagtcaccttgccca 458  
DB 361 atgtgtggaagaaagcagtgatgttgaagaaccttcagtcaccttgccca 413

RESULT 22  
LOCUS A1333667 457 bp mRNA EST 13-FEB-1999  
DEFINITION A1333667 x1 Soares, Nihmpu, SI Homo sapiens cDNA clone IMAGE:1931962  
3', mRNA sequence.

ACCESSION A1333667  
VERSION A1333667.1 GI:4070226  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 457)  
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

JOURNAL  
COMMENT

Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert length: 627 Std Error: 0.00  
Seq primer: -40up from Gibco.

FEATURES  
Source

Location/Qualifiers  
1. 457  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1931962"  
/clone\_lib="Soares Nihmpu, SI"  
/tissue\_type="Pooled human melanocyte, fetal heart, and  
pregnant uterus"  
/lab\_host="DH10B"  
/note="Organ: mixed (see below); Vector: pT73D-Pac  
(Pharmacia) with a modified polylinker; Site\_1: Not I;  
Site\_2: Eco RI; Equal amounts of plasmid DNA from three  
normalized libraries (melanocyte 2NbH, pregnant uterus  
NbHPU, and fetal heart NbH19W) were mixed, and ss circles  
were used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from pools of  
5,000 clones made from the same 3 libraries. The pools  
consisted of I.M.A.G.E. clones 260232-265223,  
340488-345479, and 484488-489479."

BASE COUNT  
ORIGIN

94 a 99 c 130 g 134 t

Query Match 39.1%; Score 411; DB 19; Length 457;  
Best Local Similarity 100.0%; Pred. No. 3.7e-216;  
Matches 411; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atgctgtgttctgtcctcctgagcgttggctggtgttcttggggagacagattagt 60  
DB 18 atgctgtgttctgtcctcctgagcgttggctggtgttcttggggagacagattagt 77  
OY 61 atataaaatccagaggtgtgttcttgcgaatggaattggtctataccagagag 120  
DB 78 atatttaattccagaggtgtgttcttgcgaatggaattggtctataccagagag 137  
OY 121 cggatccagagcgtgtcatatgttccatgtggtctctctgtgaagaagaccttcttg 180  
DB 138 cggatccagagcgtgtcatatgttccatgtggtctctctgtgaagaagaccttcttg 197  
OY 181 ccagagactcgaagtggttaacctgttcatcgtctcgtggtcagatgagtgatgg 240  
DB 198 ccagagactcgaagtggttaacctgttcatcgtctcgtggtcagatgagtgatgg 257  
OY 241 aagggagtgaaacaactggtctaccacccagagcagtgatcttctgacaccttggaga 300  
DB 258 aagggagtgaaacaactggtctaccacccagagcagtgatcttctgacaccttggaga 317  
OY 301 gcaagcttctttagtcttgaagagtggttcaatcattcactctatatttctgagaa 360  
DB 318 gcaagcttctttagtcttgaagagtggttcaatcattcactctatatttctgagaa 377  
OY 361 actcctgtgttctgagtggtcctccagtgaggaagagtgatatatgta 411  
DB 378 actcctgtgttctgagtggtcctccagtgaggaagagtgatatatgta 428

RESULT 23  
AA251015 410 bp mRNA EST 13-AUG-1997  
LOCUS z502h05.s1 NCI-CGAP-GCB1 Homo sapiens cDNA clone IMAGE:684057 3'  
DEFINITION mRNA sequence.  
ACCESSION AA251015  
VERSION AA251015.1 GI:1886114  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 410)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -41ml3 fwd. ET from Amersham  
High quality sequence stop: 402.  
FEATURES  
Location/Qualifiers  
1..410  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:684057"  
/clone\_id="NCI-CGAP-GCB1"  
/lab\_host="DH10B"  
/tissue\_type="germinal center B cell"  
/note="Vector: p7T73D-Pac (Pharmacia) with a modified  
polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA  
was prepared from human tonsillar cells enriched for  
germinal center B cells by flow sorting (CD20+, IgD-),  
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman  
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
primed with a Not I - oligo(dT) primer  
5'-GTGACCAATCTGAGAGTGGAGCGCGCCGCTCATTTTCTTTTCTTTT-3'  
1. Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified p7T73 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 94 a 88 c 108 g 120 t  
ORIGIN

Query Match 39.0%; Score 410; DB 4; Length 410;  
Best Local Similarity 100.0%; Pred. No. 1.3e-215;  
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 46 gggaagaggttgatataaataacacagagctgtgtttccgaatggaattgg 105  
|||||  
Db 1 GGGACGAGTGTATGATTAATAACACAGGGTCTGTGTTTCCAAATGGAATTGG 60  
|||||  
Oy 106 cctatacaggaagcagatccagacgtgctgcattgltccatgggtctctctgtaa 165  
|||||  
Db 61 CCTATACAGAGAGGAGTCCACAGCTGGCTGATGTCCATGGGCTTCTGTGAAA 120  
|||||  
Oy 166 gaagacattcttgccagagactgagtgtaaccttttcacgtctcctggagttac 225  
|||||  
Db 121 GAAGACCTTTTGGCCAGAGACTCGCAGTGGGTAACTTTTTCATCGTCTCGGGCTAC 180  
|||||  
Oy 226 gtcacgtgatgagtgaaggagtgacaacatggtctacacccagcagtgatcttg 285  
|||||  
Db 181 gTcATGtGATGtGAAGGAGtGAAcAAAcTGGtCTcTAcCCcAGGAGtGtCAATtTG 240  
|||||  
Oy 286 taccccttgaggaatcagctctttagcttcagcagtggttgcacatcactcc 345  
|||||  
Db 241 TACcCTTtGAGAAATcAGTtCTTtAGtCTTtGACAGtGTtGcAAATtCCATtCACTCC 300

Oy 346 ttattcttgagaaactcctgltgttcagttggtccagtgaggaagatgtat 405  
|||||  
Db 301 TTATTTTGGAGGAACCTCCTGTGTTTGGCAGTTGGCTCCAGTGAAGAGGTGAT 360  
|||||  
Oy 406 atgtagggaggaagcgaactcagtggtttgaagaccttcagtcacctgg 455  
|||||  
Db 361 ATGCTAGGAGAGGCAAACTCAGTGTGTAAGACCTTTCACTCACTTGG 410

RESULT 24  
AI022504 408 bp mRNA EST 28-AUG-1998  
LOCUS ow50e02.s1 Soares.parathyroid.tumor\_NBHPA Homo sapiens cDNA clone  
DEFINITION IMAGE:1650266 3', mRNA sequence.  
ACCESSION AI022504  
VERSION AI022504.1 GI:3237745  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 408)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
cDNA library preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo  
Ph.D.  
cDNA library arrayed by: Greg Lennon, Ph.D.  
DNA sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
[www.bio.llnl.gov/bdrp/image/image.html](http://www.bio.llnl.gov/bdrp/image/image.html)  
Insert length: 640 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham.

FEATURES  
Location/Qualifiers  
1..408  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1650266"  
/clone\_id="Soares-parathyroid.tumor\_NBHPA"  
/tissue\_type="parathyroid tumor"  
/dev\_stage="adult"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: parathyroid gland; Vector: p7T73D (Pharmacia)  
with a modified polylinker; Site.1: Not I; Site.2: Eco  
RI; 1st strand cDNA was primed with a Not I - oligo(dT)  
primer  
5'-GTGACCAATCTGAGAGTGGAGCGCGCCGCTCATTTTCTTTTCTTTT-3'  
1. Double-stranded cDNA was site selected, ligated  
to Eco RI adaptors (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of a modified p7T73  
vector (Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid  
adenomas was kindly provided by Dr. Stephen Marx, National  
Institute of Diabetes and Digestive and Kidney Diseases,  
NIH."

BASE COUNT 94 a 87 c 107 g 120 t  
ORIGIN

Query Match 38.9%; Score 408; DB 15; Length 408;  
Best Local Similarity 100.0%; Pred. No. 1.7e-214;  
Matches 408; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 46 gggaagaggttgatataaataacacagagctgtgtttccgaatggaattgg 105  
|||||  
Db 1 GGGACGAGTGTATGATTAATAACACAGGGTCTGTGTTTCCAAATGGAATTGG 60  
|||||

						(Pharmacla), digested with Not I and directionally cloned into the Not I and Hind III sites of the Latmid BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Patrina Bernaldo."
BASE COUNT	108 a	108 c	118 g	149 t	5 others	
ORIGIN						
Query Match		36.8%	Score 407;	DB 157;	Length 488;	
Best Local Similarity	100.0%;	Pred.	No. 6,1e-214;			
Matches 407;	Conservative	0;	Mismatched	0;	Indels	0; Gaps 0;
QY	40	gttttgagggaacagatttagtataatcaaccacgagcttcgttgtttlccccaatgaa	gg	99		
Dd	1	GTTTTGGGAGACGAATTAAGTATTAAATCAACOCAGSGTGCTGTGTTTTCCGAAAATGA	60			
QY	100	aattggcctatacccagaagagcgagatccccagaacgagtgcgatgtcatgtlccalgygcttctt	159			
Dd	61	AATTCGCCTAATACCGAGAAGCAGCATCCACAGCTGCCATGTCATTGCCATTTGCTCTCT	120			
QY	160	gtgaaaagaagacctttctttgccagaactcgcagcttgagtaacctgttatcatgttcccg	219			
Dd	121	GTCGAAGAAGACTTCTTTCGCACAGACTGCAGCTGGTAACCTGTTTATGTCCTCCG	180			
QY	220	gtcacgcctaigtatggtggaaggagatgaacaacatcgcgtctcacccccagaagtgtc	279			
Dd	181	CGTACCCTATGATGATGATGATGAAGGSAATGACAALACGCTCACCCCACAGAGTGT	240			
QY	280	attctgtacccctttlgsgaatgcagtlccatttcttagcttgatcagltgttgcnaatlccatt	339			
Dd	241	ATTTCGACCCCTTGGAGAACTACGTCCTTTTAGCTTGACAGTGTCCAATTCCTATT	300			
QY	340	cactccctatllcttcgagaaactcctgtgttttgcgagttgtgtcccgatgagaaga	399			
Dd	301	CATTCCTATTTTTCGAGAAGAACTCGTGTCTTTTTCAGTTGCTCCAGTGAAGAAAGA	360			
QY	400	gtgtatatgtgtagggaagcgcaactcagtgcttttgaagacctttaa	446			
Dd	361	GTGTATATGTAAGSAGAACCAACATCGTGTGAAGACTTTCACT	407			
RESULT 26						
LOCUS AA133317		566 bp	mRNA		EST	19-APR-1997
DEFINITION ESM185L213 Colon carcinoma (HCC) cell line Homo sapiens CDNA 5' end,						
ACCESSION AA133317						
VERSION AA133317.1		GI:1965646				
KEYWORDS EST.						
SOURCE human.						
ORGANISM Homo sapiens						
REFERENCE Eukaryota: Chordata: Craniata: Vertebrata: Euteleostomi;						
AUTHORS Mammalia: Eutheraia: Primates; Catarrhini: Hominoidea: Hominidae: Homo.						
I (bases 1 to 566)						
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bolt,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,C.O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai-C., Clayton,R.A., Cline,T.R., Cotton,M.B., Earle-Hughes,J.C., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M.C., Frichman,J.L., Geoghegan,N.S., Gloeck,A., Guenlu,J.C., Hanra,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marrairos,S.M., Merrick,J.M., Moreno-Palanques,R.F., McDonald,D.A., Nguyen,B.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Uitterback,T.R., Weidman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dinke,D., Feng,D.-F., Ferrite,A., Fischer,C., Hastlings,G.A., He,W.W., Huo,S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meisener,P.S., Olsen,H., Raymond L., Welty,F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Pannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.						
TITLE Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequences						



```

RESULT 28
AA164656 487 bp mRNA EST 09-MAR-1998
LOCUS
DEFINITION
2098611.r1 StrataGene ovarian cancer (#937219) Homo sapiens cDNA
ACCESSION
AA164656
VERSION
AA164656.1 GI:1740817
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 487)
AUTHORS
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Kizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,
J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wyllie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST project
TITLE
Unpublished (1997)
JOURNAL
COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1427 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 391.
Location/Qualifiers
1..487
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:594980"
/clone_lib="Stratagene ovarian cancer (#937219)"
/sex="Female"
/dev_stage="adult, 64 years"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
; Cloned unidirectionally. Primer: Oligo dT. Papillary
serous carcinoma, isolated from ascites, 64 year old
caucasian. Average insert size: 0.8 kb; Uni-ZAP XR Vector:
-5' adaptor sequence: 5' GAATTCGGCAGG 3' -3' adaptor
sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."
BASE COUNT 145 a 88 c 111 g 142 t 1 others
ORIGIN
Query Match 38.4%; Score 403; DB 3; Length 487;
Best Local Similarity 99.8%; Pred. No. 1e-211; 1; Indels 0; Gaps 0;
Matches 453; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 597 tcataagcatcagcgaagatcattcttcgtattatattcctgagctgagagttt 656
DB 1 TCATTAAGCATCTAGCAAGGATCATTCCTGATTTATTCTACGTGACGCGGTTT 60
OY 657 ggaatgaattggaagcgtataggaagacactggaacaattcagagatgctctaaagt 716
DB 61 GGATGGAATTTGGGAAGCGTTATGGGAAGACTCTGACATTTAGAGATGCTTCTAAGAT 120
OY 717 cctgttgacgctctgcaaaagtctgcagatgacatgctccttattggtggagatgc 776
DB 121 CCTTTGTTGACGCTCTGCCAAAAGTTTGCATATGCATGACATGCTTTATGCGGAATGC 180
OY 777 agtggtagagtgtaactcactgcaatgcatgacacccctcctaataagaagaagac 836
DB 181 AGTGGTAGTGTAGTCACTGTCAAGTCAATTTGACACCTCCCTATTAGGAACACAAGAC 240
OY 837 tatccttgaggcaaaacgagaagaccagcaagtcctcctaacaactctgataaagta 896
DB 241 TATCTTGGAGGCAAAACAGCAAGCAAGCAAGCAAGCTCCCTATTAACCTTGAATGA 300

```

```

OY 897 taattggaatctccggtgtttcaacatgctactcttgatataatcgccttgacct 956
DB 301 TAATTGGAATATTCCTCGGTGTTTCAACATGATCTTTGGATATATCGCTTGCTT 360
OY 957 gactgcatcaccctcttacaatatttggacaatggaaccctgatatagatcatcat 1016
DB 361 GCCTGGATATTCACCTCTTACAAATTTGGACATGATCTCGATATGATACATCAT 420
OY 1017 ttatagatgacaacccaagaattggaatgat 1050
DB 421 TTATAGATGACAAACAGAAATTCGATGAT 454

RESULT 29
AA438835 395 bp mRNA EST 14-FEB-2000
LOCUS
DEFINITION
x104g05.x1 NCI-CGAP-Ut4 Homo sapiens cDNA clone IMAGE:2778200 3',
mrna sequence.
ACCESSION
AA438835
VERSION
AA438835.1 GI:6974141
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 395)
AUTHORS
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgaps-f@mail.nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/biopr/image/image.html
Seq primer: -40UP from gibco
High quality sequence stop: 381.
Location/Qualifiers
1..395
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2778200"
/clone_lib="NCI-CGAP_Ut4"
/tissue_type="serous papillary carcinoma, high grade, 2
pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.48 kb. Life Technologies catalog #:
11542-016"
BASE COUNT 83 a 83 c 110 g 119 t
ORIGIN
Query Match 37.6%; Score 395; DB 116; Length 395;
Best Local Similarity 100.0%; Pred. No. 2.7e-207;
Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 19 ctccggcctggtggtggtgttttgggaagcagttatgatatataatcacagag 78
DB 1 CTCCTGGCGTGGTGGCGGTGTTTGGGAGACAGTTTATGATATTTAAATATCCAGGG 60
OY 79 tctgtgttttcgaaatggaatggtcctataccagagagcgatcccaagcgtgct 138
DB 61 TCTGTGTTTTCGCAAAATGCAATTTGCGCTATACAGAGAGCGGATCCAGACGTGGCT 120
OY 139 gcatgctcagggctctctctggaagaagaccttctgacagagctcagatgggt 198

```





/note="Vector: pT73D (Pharmacia) with a modified  
polylinker. Site.1: Not I; Site.2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'  
TGTTCACCATCTGGAAGTGGAGCGCCGCGATTTTCTTTTCTT 3']  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library constructed by Bento Soares and  
M. Fatima Bonaldo. RNA from normal foreskin melanocytes  
(FS374) was kindly provided by Dr. Anthony P. Albino."

BASE COUNT 120 a 85 c 101 g 138 t

ORIGIN

Query Match 37.5%; Score 394; DB 159; Length 444;  
Best Local Similarity 100.0%; Pred. No. 9,6e-207;  
Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

560 tgcagatgctacatgatttcaagcttgcgtctgtctgataagcatctagccagaagatc 619  
|||||  
39 TGCAGATGCTACATGATATTTCAAGCTTGTCTGTCTCATTAAGCATCTTACCCAGAGATC 98  
620 attctccgatttattatctacatgagcgtcgcagcgtttgagatgaattggagaagcgtatg 679  
|||||  
99 ATTCTCCGATTTATTTATTCACGTGAGCTGGCAGCTTTGGATGGAATGGAGAGCGTTTANG 158  
680 gggaagactctgaaacattcagaagatgcttcttaagatccctttgacgtctgcaaaagt 739  
|||||  
159 GGGAGACTCTGAAACATTTCCAGAGATGCTTTTAAAGATCCTTGTGACGCTGTGCAAAAGT 218  
740 ttgcagatgacatgacatgacgtcttattggtggaagtcagcgtgtagagttgacatgacatgca 799  
|||||  
219 TTTCAGATGACATGTACATGCTTTATGCTGGAGATGACAGTGAAGTAACTGACACTGCTCA 278  
800 agtcatcttgacacctccctcatataggagaagacatctcttgaagcaaaagaagcga 859  
|||||  
279 AGCATTTTGACACTCCCTCATTTAGGAAGACAGACTATCTTGAAGCAAAACAGCA 338  
860 agaacccagcgaagtcctcctaataacctgacataaagtaatttgaatattccgtgatt 919  
|||||  
339 AGAACCCAGCAAGTCCCTTAATACCTTGATATTAAGTATTTGAATATTCCTGGTTT 398  
920 tcacatgtagtacttgataaagatcgccttgcg 953  
|||||  
399 TCACATGTGTACTTTGATATGATTCGCTTGGC 432

RESULT 32  
AAB46586 392 bp mRNA EST 04-JAN-1999  
LOCUS AAB46586  
DEFINITION a37e08.s1 Soares.parathyroid.tumor\_NbHPA Homo sapiens cDNA clone  
IMAGE:1404422 3', mRNA sequence.  
ACCESSION AAB46586  
VERSION AAB46586.1 GI:2932726  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 392)  
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
JOURNAL  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaaps-r@mail.nih.gov  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo  
, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
clone distribution: NCI-CCGAP clone distribution Information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 2274 Std Error: 0.00  
Seq Primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 354.  
Location/Qualifiers  
1.392

FEATURES  
source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1404422"  
/clone.lib="Soares.parathyroid.tumor\_NbHPA"  
/tissue\_type="parathyroid tumor"  
/dev\_stage="adult"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: parathyroid gland; Vector: pT73D (Pharmacia)  
with a modified polylinker; Site.1: Not I; Site.2: Eco  
RI; 1st strand cDNA was primed with a Not I - oligo(dT)  
primer  
[5'-GGTACCAATCTGAAATGGAGCGCCGCGACCAATTTTCTTTTCTTTTCTTTT  
TTTTT-3'] double-stranded cDNA was size selected, ligated  
to Eco RI adapters (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of a modified pT73  
vector (Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid  
adenomas was kindly provided by Dr. Stephen Marx, National  
Institute of Diabetes and Digestive and Kidney Diseases,  
NIH."

BASE COUNT 92 a 81 c 105 g 114 t

ORIGIN

Query Match 37.3%; Score 392; DB 12; Length 392;  
Best Local Similarity 100.0%; Pred. No. 1.2e-205;  
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

46 gggaagcgttttgatattataatccacagcgtctgtgtgttccgaatggaattgg 105  
|||||  
1 GGAAGCGATTTGATATTTAAATCAACAGAGGCTGTGTTTCCGAATGGAATTTGG 60  
106 cctataccagagagcgaatccacagcgtgctgcatgttccatgaggtctctctgtgaaa 165  
|||||  
61 CCTATACAGAGAGAGGAGGATCCAGAGCTGGCTGATGTCCATGAGGCTCTCTGTGAAA 120  
166 gaaagccttcttgccagagactgcagtgtaacctgtttcatctctctctggtacc 225  
|||||  
121 GAAGACCTTTCTTGCCAGAGACTGCGAGTGGTAACCTGTTTCAATCTCTGGGCTACC 180  
226 gtcacgtgacatgctgaaagagagtgaaacaaactgcttaccacccagcagtgatcttgc 285  
|||||  
181 GTCATGTGATGTGGAAGGAGAGTGAACAAACTGCTTACCCGAGCAGTGCATTTGC 240  
286 tacccttggagagatcagatcttctttagcttgagagagtgtaaaatctatcacc 345  
|||||  
241 TACCCCTTGGAGATGCAAGTCTTTTAGTCTTGACAGGTGGCAAAATTCATTCATCTCC 300  
346 ttatttctgaggaactcctgtgttcttgacgttgcctccagtgaggaagatgcat 405  
|||||  
301 TTATTTTCTGAGCAAACTCTGTGTTTGCAGTGTGCTCCAGTGAAGAAAGTGTAT 360  
406 atggtgaggaagcgaactcagtgatttgaaga 437  
|||||  
361 ATGCTAGGAGAGCAACTCACTGTGTTGAAGA 392

RESULT 33  
A1028502 427 bp mRNA EST 28-AUG-1998  
LOCUS A1028502  
DEFINITION ow44a03.x1 Soares.parathyroid.tumor\_NbHPA Homo sapiens cDNA clone  
IMAGE:169644 3', mRNA sequence.  
ACCESSION A1028502  
VERSION A1028502.1 GI:3245811  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
1 (bases 1 to 427)  
AUTHORS  
TITLE  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)

JOURNAL  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-f@mail.nih.gov  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo  
, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.dio.llnl.gov/dbfp/image/image.html  
Insert Length: 616 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 413.  
Location/Qualifiers

## FEATURES

source

1. 427  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1649644"  
/clone\_lib="Soares\_Parathyroid\_tumor\_NbHPA"  
/tissue\_type="Parathyroid tumor"  
/dev\_stage="adult"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: parathyroid gland; Vector: pT73D (Pharmacia)  
(with a modified polylinker; Site 1: Not I; Site 2: Eco  
RI; 1st strand cDNA was primed with a Not I - oligo(dT)  
primer  
15'-TGTTGCCAATCTGAAGTGGAGCGCCGCCAATTTTCTTTTCTTTT  
TTTCT-3', double-stranded cDNA was size selected, ligated  
to Eco RI adapters (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of a modified pT73  
vector (Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid  
adenomas was kindly provided by Dr. Stephen Marx, National  
Institute of Diabetes and Digestive and Kidney Diseases,  
NIH."

BASE COUNT 97 a 96 c 110 g 124 t  
ORIGIN

Query Match 37.3%; Score 392; DB 15; Length 427;  
Best Local Similarity 100.0%; Pred. No. 1.2e-205;  
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 48 gaacgagttatataataaccacagagctcgtgtgtttccgaatggaattgac 107  
|||||  
Db 1 GAACGAGTTTATATTAATAATCACCAGGCTGTGTGTTTCCGAATGCAATGGCC 60  
OY 108 tatacagaagaagagatccacagctgctcatttcacatgagcttcgtgaaaga 167  
|||||  
Db 61 TATACAGAGAGGCGGATCCAGACGCGCTCATTTGTCATGGCGTCTCTGTAAGA 120  
OY 168 agaccttcttggcagagactcgagtyggtaacctgttcaatcgtccctggacac 227  
|||||  
Db 121 AGACCTTCTTGGCAGACGACGACGTGGGTAACTGTTTCACTGCTCCGGCTACCGT 180  
OY 228 catggtgatggtgaaggaagtgaacaaactgcttaccacccagagagtgatattgta 287  
|||||  
Db 181 CATGCTGATGGAAGAGGAGTGAACAACTGGCTTACCCCGACAGGTGTATTTGCTA 240  
OY 288 ccccttggagaatgagcttcctttgaatcagagtggtgaaatcattcactcctt 347  
|||||  
Db 241 CCCTTGGAAATGCACTTCCTTTAGTCTGACAGTGTTCGAAATTCATTCACCTCTT 300  
OY 348 atttcggaagaactcctgtgtgtttgagagtgctccacagtggaagaagattat 407  
|||||  
Db 301 ATTTTCTGAGAAACTCCTGTGTTGTTTGCAGTTGGCTCCGAGTGAAGAAAGGTATAT 360

OY 408 ggtagggaaggcaactcagtggttgaagacc 439  
|||||  
Db 361 GGTAGGGAAGGCAAACTCAGTGTGGAAGACC 392

## RESULT 34

LOCUS BG538574 878 bp mRNA EST 03-APR-2001  
DEFINITION 602567303P1 NIH\_MGC\_77 Homo sapiens CDNA clone IMAGE:4691712 5',  
mRNA sequence.

ACCESSION BG538574  
VERSION BG538574.1 GI:13530807  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
1 (bases 1 to 878)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-f@mail.nih.gov  
Tissue Procurement: CLONTECH Laboratories, Inc.  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
plate: L1CM1511 row: f column: 01  
High quality sequence stop: 615.  
Location/Qualifiers

## FEATURES

source

1. 878  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4691712"  
/clone\_lib="NIH MGC\_77"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: Lung; Vector: pDNR-LIB (Clontech); Site 1:  
SfiI (ggcgccctcgcc); Site 2: SfiI (ggcctatggcc); 5' and  
3' adaptors were used in cloning as follows: 5' adaptor  
sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence:  
5'-ATTCTAGAGCGCCAGCGCGGACATG-dT(30)BN-3' (where B = A,  
C, or G and N = A, C, G, or T). Average insert size 1.9  
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts  
by PCR. This library was enriched for full-length clones  
and was constructed by Clontech Laboratories (Palo Alto,  
CA). Note: this is a NIH\_MGC Library."

BASE COUNT 278 a 121 c 200 g 279 t  
ORIGIN

Query Match 36.5%; Score 383; DB 15; Length 878;  
Best Local Similarity 100.0%; Pred. No. 1.3e-200;  
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 668 ggaagcgtatggggaagactcgaacaaatcagaagatgcttcaagatccttgtaag 727  
|||||  
Db 1 GGAAGCGTATTGGGGAAGACCTGGAACAAATTCAGAGATGCTTAAAGATCCTTGTGAGC 60  
OY 728 ctctgcaaaagcttcagatgaatctacagctcttatggtggagtcagtgtaagt 787  
|||||  
Db 61 CTCGCAAAAGTTTCAGATGACATGTACAGTCTTATAGTGGGAATGCAAGTGAAGT 120  
OY 788 taatcagtcagatcatttgaacacctccctcatlaaggaacaagactatccttgagg 847  
|||||  
Db 121 TAGTCATGTCGAACGATTTTACACCTCCCTCATTTAGAGAACAGACATATCTTTGAGG 180  
OY 848 caaacaagcgaagaaccagcaagtcacctataacattgacataaagtaatttgaat 907  
|||||  
Db 181 CAAACAGCAGGAAGAACCCAGACAGTCCCTATTAACCTTGCAATATAGTATTTTGAT 240

```

OY 908 atccgtgatttcaacatggtacttgatgataatgatccgttcgttcgtgattgata 967
    |||
DB 241 ATTCCGGTGGTTTCAACATGCTGTAATTGGATATATCATGCGCTTGCGCTGCGATGA 300
    |||
OY 968 tcaaccttacaatatttggaaatggatccctgatalgatagatcaatcattatagatga 1027
    |||
DB 301 TCACCTCTTACATATTTTGGACATGATCTCGATATGATGACATCATTTATAGATGA 360
    |||
OY 1028 caaacagaagatcgaaatgat 1050
    |||
DB 361 CAAACCAAGATTCGATGAT 383
    |||

RESULT 35
AA584287 433 bp mRNA EST 08-SEP-1997
LOCUS nm75h07.s1 NCI_CGAP_C09 Homo sapiens cDNA clone IMAGE:1089757 3',
DEFINITION mRNA sequence.
ACCESSION AA584287
VERSION AA584287.1 GI:2368896
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 433)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 364.
Location/Qualifiers
1. 433
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1089757"
/clone_lib="NCI_CGAP_C09"
/tissue_type="Colon tumor RER+"
/lab_host="DH10B"
/notes="Organ: Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
RER+ colon tumor, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT7T3
vector. Library is not normalized. Library was
constructed by Bento Soares and M. Fatima Bonaldo (Soares4
) ."
BASE COUNT 93 a 90 c 122 g 128 t
ORIGIN

```

```

Query Match 35.8%; Score 376; DB 9; Length 433;
Best Local Similarity 100.0%; Pred. No. 9e-197;
Matches 376; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 45 ggggaagagatttgatataaatacagaggtctgtatttcgaaatggaatg 104
    |||
DB 33 GGGGACCAAGATTGTGATATTAATACACAGGCTCTGTGTTCCGAAATGGAATTG 92
    |||
OY 105 gacctaccagagagcgcgacgcagctggtcgtcatgtccatgggcttctctgtga 164
    |||

```

```

DB 93 GCCATACACAGAGAGCGCATCCCAACACTGCTGCATTTGTCATGAGGCTTCTGTGAA 152
    |||
OY 165 agaagaccttcttggccaggactcgagtggtgaacctgttccatgcctcctgggtac 224
    |||
DB 153 AGAAGACCTTCTTGGCCAGGACTCCGACTGGGTAACTGTTTCATCCGCTCGGGCTAC 212
    |||
OY 225 cgtcagtgtgagtgaaggagatgaacaactgctctacccccaggcaatgtcatttc 284
    |||
DB 213 CGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 272
    |||
OY 285 gtacaccttggagaatcgatctcctttagtcttgacagtggtgaatccatcactc 344
    |||
DB 273 GTACCCCTTGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 332
    |||
OY 345 ctatttcttgagaactcctgtgtgttgcagtggtgctcccaatgagaagaatgata 404
    |||
DB 333 CTATTTTCTGAGGAACACTCTGTTGTTTGCATTTGCTCCCATGAGGAAGATGTA 392
    |||
OY 405 tatgtaagggaagca 420
    |||
DB 393 TATGCTAGGGAAGCA 408
    |||

RESULT 36
AA903999 433 bp mRNA EST 21-APR-1998
LOCUS oe74a02.s1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1417322 3',
DEFINITION mRNA sequence.
ACCESSION AA903999
VERSION AA903999.1 GI:3039122
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 433)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html
Insert length: 623 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham.
Location/Qualifiers
1. 433
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1417322"
/clone_lib="NCI_CGAP_Lu5"
/tissue_type="carcinoid"
/lab_host="DH10B"
/notes="Organ: Lung; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
neuroendocrine lung carcinoid, and was then primed with a
Not I - oligo(dT) primer. Double-stranded cDNA was ligated
to Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT7T3 vector. Library is normalized. Library was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 101 a 93 c 112 g 126 t 1 others
ORIGIN

```

Query Match 35.8%; Score 376; DB 13; Length 433;

Best Local Similarity 100.0%; Pred. No. 9e-197;

Matches 376; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

46 gggagagagtagtagatataataacacagaggtcgtgtgttccgaaatgaaatg 105

1 gggagagagtagtagatataataacacagaggtcgtgtgttccgaaatgaaatg 60

106 cctatccag 165

61 cctatccag 120

166 gaaagacaccttctgagcagagagagagagagagagagagagagagagagag 225

121 gaaagacaccttctgagcagagagagagagagagagagagagagagagagag 180

226 gtaatgtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 285

181 gtaatgtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 240

286 taaccttctgag 345

241 taaccttctgag 300

346 ttatttctgag 405

301 ttatttctgag 360

406 atgtaggag 421

361 atgtaggag 376

# RESULT 37

AA313617

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

Other ESTs: THC186857

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

tel: 3018699056

Fax: 3018699423

Email: arkerlavage@tigr.org

For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hg1/hg1.html>)

Seq primer: M13 Reverse.

## FEATURES

SOURCE

1. 417

/organism="Homo sapiens"

/db\_xref="ATCC (Inhost):109742"

/clone\_lib="Colon carcinoma (HCC) cell line"

/tissue\_type="colon"

/cell\_line="KM12C"

/note="Organ: colon; Vector: pBluescript SK-; Site:1; Ecor1; Site:2; Xho1"

BASE COUNT 107 a 92 c 92 g 126 t

ORIGIN

Query Match 35.7%; Score 375; DB 5; Length 417;

Best Local Similarity 100.0%; Pred. No. 3.2e-196;

Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

460 ctccgaatgcctcgttctcaagaagaactcgttctcagttcactcccccattctcgt 519

43 ctccgaatgcctcgttctcaagaagaactcgttctcagttcactcccccattctcgt 102

520 agtagaacaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 579

103 agtagaacaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 162

580 tcaagctctgctctcgtcgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 639

163 tcaagctctgctctcgtcgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 222

640 ctgagacctgag 699

223 ctgagacctgag 282

700 agagatgcttcaagaatccttctgagcagcctcgtcgaagaatttgcagatgacatgacat 759

283 agagatgcttcaagaatccttctgagcagcctcgtcgaagaatttgcagatgacatgacat 342

760 cttaatgag 819

343 cttaatgag 402

820 attagaagaagaagaag 834

403 attagaagaagaagaag 417

## RESULT 38

BE312617

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

COMMENT

Other ESTs: NIH\_MGC\_19

Contact: Robert Strausberg, Ph.D.

Email: cgaabs-remail.nih.gov









source

1. .496  
/organism="Homo sapiens"  
/db\_xref="GDB:5924822"  
/db\_xref="taxon:9606"  
/clone="IMAGE:727910"  
/clone\_lib="Soares\_testis\_NHT"  
/sex="male"  
/lab\_host="PH10B"  
/note="Vector: pT773D-Pac (Pharmacia) with a modified polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTGGAGCGGCCGCGCATTTTCTTTTCTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT773 vector. Library went through one round of normalization to Cots, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 117 a 112 c 123 g 144 t

ORIGIN

Query Match 34.6%; Score 363; DB 7; Length 496;  
Best Local Similarity 99.8%; Pred. No. 1.4e-189;  
Matches 413; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 46 gggagcagattatataataaacacacagagctgctgttccggaatggaattg 105  
|||||  
DB 1 GGGAGCAGATTATATATTAATACACAGGCTGTGTTCCGAATGGAATGG 60  
OY 106 ccataccgagagagcgagctccagagctgctgcatgtgctgagctctctgtgaa 165  
|||||  
DB 61 CCTATACCGAGAGCGGATCCAGACGGCTGCATTGTCATGGGCTTCGTGAAA 120  
OY 166 gaagaccttcttgccagagctcgagctggttaacctgttccatgctccctggtacc 225  
|||||  
DB 121 GAAGACCTTTCTTGCCAGCTCGAGTGGGTAACTGTTTCATGCTCCGCGCTACC 180  
OY 226 gtaatgtagatgtagaagagtgaaacactgcttaccctcccgaggaagatattg 285  
|||||  
DB 181 GTCATGTCATGTCGAAGAGAGTGAACAACTGGCTCTACCCCGAGGAGTGTATTCG 240  
OY 286 tacccttggaagatgagcttctttagcttgacagtggttgcaaatcattaccctc 345  
|||||  
DB 241 TACCTTTGGAGAAATGCAAGTTCCTTTAGTCTGACAGTGTTCAAATTCATTCACTGC 300  
OY 346 ttattcttgaggaactcctgtgtgttttgagttgctcccgagtgaggaagatgtat 405  
|||||  
DB 301 TTATTTTTCGAGAAACTCTGTGTTGTTTTCAGTTGGCTCCAGTGAAGAGATGTAT 360  
OY 406 atggttaggaaggaacactcagtggttggaacacttcaagtcaccttgagcag 459  
|||||  
DB 361 ATGTGAGGAGCAAACTCAGTGTTCGAAGACCTTCACTGACCTCGCGCAG 414

RESULT 44

AA713478 460 bp mRNA EST 29-DEC-1997  
LOCUS nv67a03.s1 NCL\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:1234828 3'  
DEFINITION mRNA sequence.  
ACCESSION AA713478  
VERSION AA713478.1 GI:2725752  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 460)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.

source

Email: cgaps-remail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,  
Ph.D., Gerald Marti, M.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
found through the I.M.A.G.E. Consortium/LNLW at:  
[www.bio.linnl.gov/birop/image/image.html](http://www.bio.linnl.gov/birop/image/image.html)  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 433.  
Location/Qualifiers

1. .460  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1234828"  
/clone\_lib="NCL\_CGAP\_GCB1"  
/tissue\_type="germinal center B cell"  
/lab\_host="PH10B"  
/note="Vector: pT773D-Pac (Pharmacia) with a modified polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTGGAGCGGCCGCGCATTTTCTTTTCTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT773 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 107 a 106 c 115 g 132 t

ORIGIN

Query Match 34.5%; Score 362; DB 11; Length 460;  
Best Local Similarity 99.8%; Pred. No. 5.1e-189;  
Matches 412; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 46 gggagcagattatataataaacacacagagctgctgttccggaatggaattg 105  
|||||  
DB 1 GGGAGCAGATTATATATTAATACACAGGCTGTGTTCCGAATGGAATGG 60  
OY 106 ccataccgagagagcgagctccagagctgctgcatgtgctgagctctctgtgaa 165  
|||||  
DB 61 CCTATACCGAGAGCGGATCCAGACGGCTGCATTGTCATGGGCTTCGTGAAA 120  
OY 166 gaagaccttcttgccagagctcgagctggttaacctgttccatgctccctggtacc 225  
|||||  
DB 121 GAAGACCTTTCTTGCCAGCTCGAGTGGGTAACTGTTTCATGCTCCGCGCTACC 180  
OY 226 gtaatgtagatgtagaagagtgaaacactgcttaccctcccgaggaagatattg 285  
|||||  
DB 181 GTCATGTCATGTCGAAGAGAGTGAACAACTGGCTCTACCCCGAGGAGTGTATTCG 240  
OY 286 tacccttggaagatgagcttctttagcttgacagtggttgcaaatcattaccctc 345  
|||||  
DB 241 TACCTTTGGAGAAATGCAAGTTCCTTTAGTCTGACAGTGTTCAAATTCATTCACTGC 300  
OY 346 ttattcttgaggaactcctgtgtgttttgagttgctcccgagtgaggaagatgtat 405  
|||||  
DB 301 TTATTTTTCGAGAAACTCTGTGTTTTCAGTTGGCTCCAGTGAAGAGATGTAT 360  
OY 406 atggttaggaaggaacactcagtggttggaacacttcaagtcaccttgagcag 458  
|||||  
DB 361 ATGTGAGGAGCAAACTCAGTGTTCGAAGACCTTCACTGACCTCGCGCA 413

RESULT 45

AI682889 393 bp mRNA EST 17-DEC-1999  
LOCUS AI682889

DEFINITION	REFERENCE	JOURNAL	COMMENT	FEATURES
NCBI CGAP_Pan1 Homo sapiens cDNA clone IMAGE:232901 3'	MRNA sequence.			
ACCESSION	U669703.x1			
VERSION	1			
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
	1 (bases 1 to 393)			
	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .			
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),			
	Tumor Gene Index			
	Unpublished (1997)			
	Contact: Robert Strausberg, Ph.D.			
	Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a>			
	Life Technologies catalog #: 11548-013			
	DNA sequencing by: Washington University Genome Sequencing Center			
	clone distribution: NCI-CGAP clone distribution information can be			
	found through the I.M.A.G.E. Consortium/BLM at:			
	<a href="http://www.bio.lnln.gov/bdrp/image/image.html">www.bio.lnln.gov/bdrp/image/image.html</a>			
	Insert length: 639			
	Seq primer: -400P from Gidco			
	High quality sequence stop: 370.			
	Location/Qualifiers			
	1..393			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone_1b="NCI-CGAP_Pan1"			
	/tissue_type="adenoecarcinoma"			
	/lab_host="DH10B"			
	/note="Organ: pancreas; Vector: pCMV-SPORT6; Site: 1; SalI;			
	/site-2: NotI; Cloned unidirectionally. Primer: Oligo dT.			
	Average insert size 1.72 kb. Life Technologies catalog #:			
	11548-013"			
	83 a 83 c 108 g 118 t 1 others			
	BASE COUNT			
	ORIGIN			
	33.7%; Score 354; DB 23; Length 393;			
	Best local similarity 100.0%; Pred. No. 1.4e-184;			
	Matches 354; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
	Query Match			
	19 cctctgacgttggtagcgggtgtgttggggaacagtttagtataataaaccacaagg 78			
	Db 1 cttccctgcggttggtgctgctgctggaagaaacgtttgagttatgattaaatcaccacgg 60			
	79 tctgtgttttcgcgaatcgaatctgacctatcacgaagaaacgagatccacagcgttgct 138			
	Db 61 tctgtgtgttttcgcgaatcgaatgagaaatgagctatcacgaagaaacgagatccacagcgttgct 120			
	139 gattatgacatggcttctctctggaagaagaaccttcttggccaggaaccgcagatggct 198			
	Db 121 gattatgacatggcttctctctggaagaagaaccttcttggccaggaaccgcagatggct 180			
	199 aacctgttcatcgttcctcccggtcaccgacgtcgtcgtgtagtgaagagagtgaaacaactg 258			
	Db 181 AACCTGTTTCATGTCCTCGGGCACCGTCATGATGTGTAAGAGGAGTGACCAACACG 240			
	259 gctctacccccagagatgatttctgtacaccttggagaatgaagcgttctttagtctt 318			
	Db 241 GCTCTACCCCCAGAGATGATTCTGTACACCTTGGAGAATGACGTTCTTTAGTCCTT 300			
	319 gacagtgctgcaaatcattcactctctatttcttctgaggaacactcctgtgtct 372			
	Db 301 GACAGTGTTCGCAAAATTCATTCACCTCTTATTTCGAGAGAAATCCGTTGTT 354			
DEFINITION	RC6-HT0594-180500-011-H03 HT0594 Homo sapiens cDNA, mRNA sequence.			
LOCUS	BE709243			
RESULT	46			
	BE709243	462 bp	mrna	EST
				12-SEP-2000

ACCESSION	BE709243
VERSION	BE709243.1
KEYWORDS	GI:10097508
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Mukaryoty; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
	1 (bases 1 to 462)
	Dias Neto, E., Garcia Correa, R., Vertovski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bial, G. S., Simpson, D. H., Brunshtein, A., deOliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, J. M., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J.
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE	200202663
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?PL=6t2-Cg-Rc6-HT0594-180500-011-H036t3-2000-05-18&t4-1) Seq primer: puc 18 forward High quality sequence start: 18 High quality sequence stop: 346. Location/Qualifiers 1..462 /organism="Homo sapiens" /rd_xref=taxon:9606 /clone_lib="HT0594" /dev_stage="Adult" /note="Organ: head,neck; Vector: puc18; Site:1; Smar: Site_2; Smar: A mini-library was made by cloning products derived from ORSTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT	136 a 84 c 105 g 137 t
ORIGIN	
Query Match	33.7%; Score 354; DB 139; Length 462;
Best Local Similarity	100.0%; Pred No. 1,4e-184;
Matches 354;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y	684 agactctgaacaatcgaagaagctctctaaatcctctgtgacgctctgcaaaagtgc 743
Db	87 AGACTCTGAACAATTCAGAGATGCTCTTCAAGATCTCTGTTGAGCGCTCTGCAAAAGTTTGC 146
Y	744 agatgaacatgaacgtctcttatgtgtggaaagcagcggtgaagtatgaactcgttaactc 803
Db	147 AGATGACATGTACAGTCTTATATGTGTGGAAATGCAGTGTAGAGTTATCATCTACTGTCAATC 206
Y	804 atttgacacccctccatctaggaagaagactatcctctgaggcaaaagaagaagaa 863
Db	207 ATTGAGACCTCCCTCATTTAGGAGACGAAGACATCTCTTGAGCGCAAAACGAAGCCAGAA 266
Y	864 ccagcgaagccctataacttgcataaagatataatttgaaatattcgttgatttcaa 923
Db	267 CCAGCAAGACGCTTATACCTCTGCATATTAATATTTTGAATATTCGTGTGTTTCAA 326
Y	924 catgtactcttgataatgatgcgcttgacctgtgcttgatatacactcttaacatat 983
Db	327 CATGTACTCTTGATATATGATCCCTTGCGCTTGCGTGTGATTTTCACCTCTTAATAT 386

QY 984 ttgagacatgacatcctgagatagatagatcattatagagatcacaccagaa 1037  
|||||  
DB 387 TTGGACATGATGATCTCGATATATAGATCATCTTTATAGATGACAAACCA 440

RESULT 47  
AA576267 462 bp mRNA EST 09-SEP-1997  
LOCUS AA576267/c  
DEFINITION nm62d08.s1 NCI\_CGAP\_Br3 Homo sapiens cDNA clone IMAGE:1072815 3',  
mRNA sequence.  
ACCESSION AA576267  
VERSION AA576267.1 GI:2350782  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
1 (bases 1 to 462)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue procurement: L. Jeffrey Medeiros, M.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Stratagene, Inc.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)  
Insert Length: 1377 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 431.  
Location/Qualifiers  
SOURCE  
1. 462  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1072815"  
/clone\_1ib="NCI\_CGAP\_Br3"  
/sex="Female"  
/tissue\_type="breast tumor"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: breast; Vector: Bluescript SK-; Site: 1:  
EcoRI; site: 2: XhoI; Cloned unidirectionally. Primer:  
Oligo dT. Ductal breast tumor. 5' adaptor sequence: 5'  
GAATTCGACGACG 3' 3' adaptor sequence: 5'  
CTCAGCTTTTCTTTTCTTTTCTTTT 3' Average insert size: 0.9 kb."  
BASE COUNT 149 a 99 c 76 g 138 t  
ORIGIN

Query Match 33.6%; Score 353; DB 9; Length 462;  
Best Local Similarity 100.0%; Pred. No. 5e-184;  
Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 698 tcagagatgcttctaagatcctgttgacgctcgtgcaaaagtttcagatgacatgaca 757  
|||||  
DB 462 TCAGAGATGCTCTTAAGATCCTGTGAGCGCTGCAAAAGTTTCAGATGACATGTACA 403

QY 758 gtcttatagtggaatgagatgtagatgtagtcatgctgcaagtcatttgaacccccc 817  
|||||  
DB 402 GTCTTTATGTTGGGAATGCAAGTGTAGACTTGTACTGCTCAAGTATTTTGACACTGCC 343

QY 818 tcattgagaagaagaactatccttgaggcaaaacaagcgaaagaccagcaagtcct 877  
|||||  
DB 342 TCATTGGAAGCAAGAGACTATCTTGAAGCAAAAGCAAGAGAACCCAGAGTCCCT 283

QY 878 ataacctgcatatagataatttgaatattcgtggttttcaacatgtaacttgg 937  
|||||  
DB 282 ATAACTGCTCATATATAGATATTTTGAATATTCGTCGTTTCAACATGTACTTGG 223

QY 938 taatgtacgcttgagccttgctgctgtattatcactcttacaatttggaaatgacatc 997  
|||||  
DB 222 TAATGATCGCCTTGAGCTGTGCTGTATATACCTCTTCAAAATTTGGAAATGATGC 163

QY 998 ctgagatagatagacatcattatagatgacaaacagaagatcgaatggat 1050  
|||||  
DB 162 CTGATATATGATGATCATCTTTATAGATGACAAACGAAAGATTGCAATGATGAT 110

RESULT 48  
AA255918 456 bp mRNA EST 13-AUG-1997  
LOCUS AA255918/c  
DEFINITION zs27g02.t1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:686450 5',  
mRNA sequence.  
ACCESSION AA255918  
VERSION AA255918.1 GI:1891486  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
1 (bases 1 to 456)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 627 Std Error: 0.00  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 433.  
Location/Qualifiers  
SOURCE  
1. 456  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:686450"  
/clone\_1ib="NCI\_CGAP\_GCB1"  
/tissue\_type="germinal center B cell"  
/lab\_host="DH10B"  
/note="Vector: pRT30-Pac (Pharmacia) with a modified  
polylinker; Site: 1: Not I; Site: 2: Eco RI; 1st strand cDNA  
was prepared from human tonsillar cells enriched for  
germinal center B cells by flow sorting (CD20+, IgD-),  
provided by Dr. Louis M. Staudt (NCI). Dr. David Allman  
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
primed with a Not I - oligo(dT) primer  
[5'-TGTTACCAATCTGAAGTGGAGCGCGCTCATTTTCTTTTCTTTT-3'  
]. Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pRT30 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 146 a 98 c 77 g 135 t  
ORIGIN

Query Match 33.1%; Score 348; DB 4; Length 456;  
Best Local Similarity 100.0%; Pred. No. 2.9e-181;  
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 703 gatgcttcaagatcctgttgacgctcgtgcaaaagtttcagatgacatgacatgac 762  
|||||  
DB 455 GATGCTTCTTAAGATCCTGTGAGCGCTGCAAAAGTTTCAGATGACATGTACAGCTT 396

QY 763 tatgtgggaagacgtgtagatgtagtcaatgcatgacatgacatgacatgacatgac 822  
|||||  
DB 395 TATGTTGGGAATGCAAGTGTAGACTTGTACTGCTCAAGTATTTTGACACTGCCCTCAT 336

QY 823 aggaagaagaagactatccttgaggcaaaacaagcgaaagaccagcaagtcctataac 882  
|||||

Db	335	AGGAAGCAAGGACATCTCTTGAGGCAAAACAAAGCAAGCAAGCAAGCCCTTATAC	276
Qy	883	cttgcataatagtataatttggatatattccgtggtttcaaatgtaactcttggataatg	942
Db	275	CTTGGAATATAGTATATATTTTGAATATCCGAGTCTTTCACATGGTACTTTGGATTAATG	216
Qy	943	atgcgccttggcccttggcgcgttattatcacctcttaacaatttggaaatagatccttga	1002
Db	215	ATCGCGCTTGGCCCTTGGCTGTGATTATTCACCTCTTACAAATATTTGGAGAACATGATCTCGGA	156
Qy	1003	tatgtagcatcatatagatgatacaacaacgaagaattcgaattgatc	1050
Db	155	TATGATAGCATCTTTATATAGATGATACAAACGAGAAATGTTGAATTGAT	108

RESULT	49
N35735	
LOCUS	N35735 431 bp mRNA EST 16-JAN-1996
DEFINITION	Yx81ad06.r1 Soares melanocyte 2NbHM Homo sapiens cDNA clone IMAGE:26814 5' , mRNA sequence.
ACCESSION	N35735
VERSION	N35735.1 GI:115687
KEYWORDS	EST,
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE AUTHORS	1 (bases 1 to 431)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,	

TITLE	The Washu-Merck EST Project
JOURNAL	Unpublished (1995)
COMMENT	Contact: Wilson RK

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [estewatson.wustl.edu](mailto:estewatson.wustl.edu)  
High quality sequence stops: 325  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert length: 1351      Std Error: 0.00  
Seq primer: 17  
High quality sequence stop: 325.

FEATURES	Location/qualifiers
source	1, .431

```

/organism="Homo sapiens"
/db_xref="GDB:3877756"
/db_xref="taxon:9606"
/clone="IMAGE:268114"
/clone_lib="Soares melanocyte 2NBHM"
/sex="Male"

```

```

"/lab_host="DHI0B (ampicillin resistant)"
/Note="Vector: pT73D (Pharmacia) with a modified
polylinker. Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer. 15'
TGTTACCAATCTGAAGTGGAGCGCGCCATCTTTTTTTTTTTTTTTT 3'}.
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73D vector
(Pharmacia). Library constructed by Bento Soares and
M. Felicia Bonalco. RNA from normal foreskin melanocytes
(F5374) was kindly provided by Dr. Anthony P. Albino."
BASE COUNT      116 a      94 c      94 g      124 t      3 others
ORIGIN

```

Query Match	33.0%;	Score 346;	DB 159;	Length 431;
Best Local Similarity	100.0%;	Pred. No. 3.7e-180;		

Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY	464	ggttcgcgctggtttaaaaaaacctcgtcttcgaagtaacccccccaattctctagta	523
Db	1	gtaatccgcctgtttcaagaaacctcttcttgaattcctccctcaattctctagta	60
OY	524	ggaacaaatgaagttgcagctgcctcttcttcgaactgcagtgcaatgataattcaa	583
Db	61	ggacacatgaaagttgacccgcctcttcttgcgaactgcgaagtgcatcatgattttcaa	120
OY	584	gcttcgcgtctgcataagcatctgcgaagaatcattctccgtatattatcactga	643
Db	121	gccttcgctgcctcctctatgaacatctacccaagacatctctcctgattttatttcactgg	180
OY	644	agctgcgagcttctggaatattggaagcgttatggggaagctctgaacaatcagag	703
Db	181	acctgcgacgctgttgatataaattgggaacgctttatggggaagccttgacaaatttcagag	240
OY	704	atgcctctaagatcctcttbaagcctctgcaaaaatttgacagtgacatgacagctctt	763
Db	241	atgcctctaaagaccctgtgtgacgcctctgcamaatttgacagatgacatgacgctttt	300
OY	764	atgatgggaatcagatggtgataagttatgacatgcgtcaagtaatttga	809
Db	301	atggctgggaattgcagtggtgaagttatgtagtactgtaacgttaattttga	346

RESULT 50

LOCUS	A1128376	410 bp	mRNA	EST	24-SEP-1998
DEFINITION	ox86c12.s1 Soares.senescent_fibroblasts.Nbhsf Homo sapiens cDNA clone IMAGE:1663222 3', mRNA sequence.				

VERSION	AI128976.1	GI:3597490
KEYWORDS	EST.	
SOURCE	human.	

REFERENCE  
1 (bases 1 to 410)  
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

Tumor Gene Index  
JOURNAL  
Unpublished (1997)  
COMMENT  
Contact: Robert Strausberg, Ph.D.

This clone is available royalty-free through LNL ; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 583 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 388.

## FEATURES

## SOURCE

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="1663222"
/clone_lib="Scars senescent fibroblasts_NbHSF"
/tissue_type="senescent fibroblast"
/lab_host="DH10B (ampicillin resistant)"
/notes="Vector: pT73D (Pharmacia) with a modified
polylinker V_type: phagemid; Site:1: Not I; Site:2: Eco RI
; 1st strand cDNA was primed with 1: Not I; oligo(dn)
primer [5',
TGTACCAATCTGAAGTGGAGCGGCCCATTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(normalization). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento

```

BASE COUNT	94 a	87 c	109 g	120 t
ORIGIN				

Query Match 32.6%; Score 342; DB 16; Length 410;  
Best Local Similarity 99.7%; Pred. No. 6, 1e-178;  
Matches 392; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 46 gggagcgaatttaagatataataataaccagggtctgtgttccgaaatgaaattgg 105  
|||||  
Db 1 gggacgcaattttatgattttaaataaccagggctgtgttccgaaatgaaattgg 60  
|||||

QY 106 cctataccagagagcagatccagacgtgtgtgattgtccalgggcttctgtgaaa 165  
|||||  
Db 61 CCTATACGAGAGAGCGGATCCAGACGCGCTGCTATGTCATGGGCTTCTCTGTGAAA 120  
|||||

QY 166 gaagaccttcttggccaagactcgcagtggtaacctgttcatcgtctcggctacc 225  
|||||  
Db 121 GAAGACCTTCTTGCCAGGACTCGCAGTGGGTAACTGTTTCATGTCCTCGGGCTACC 180  
|||||

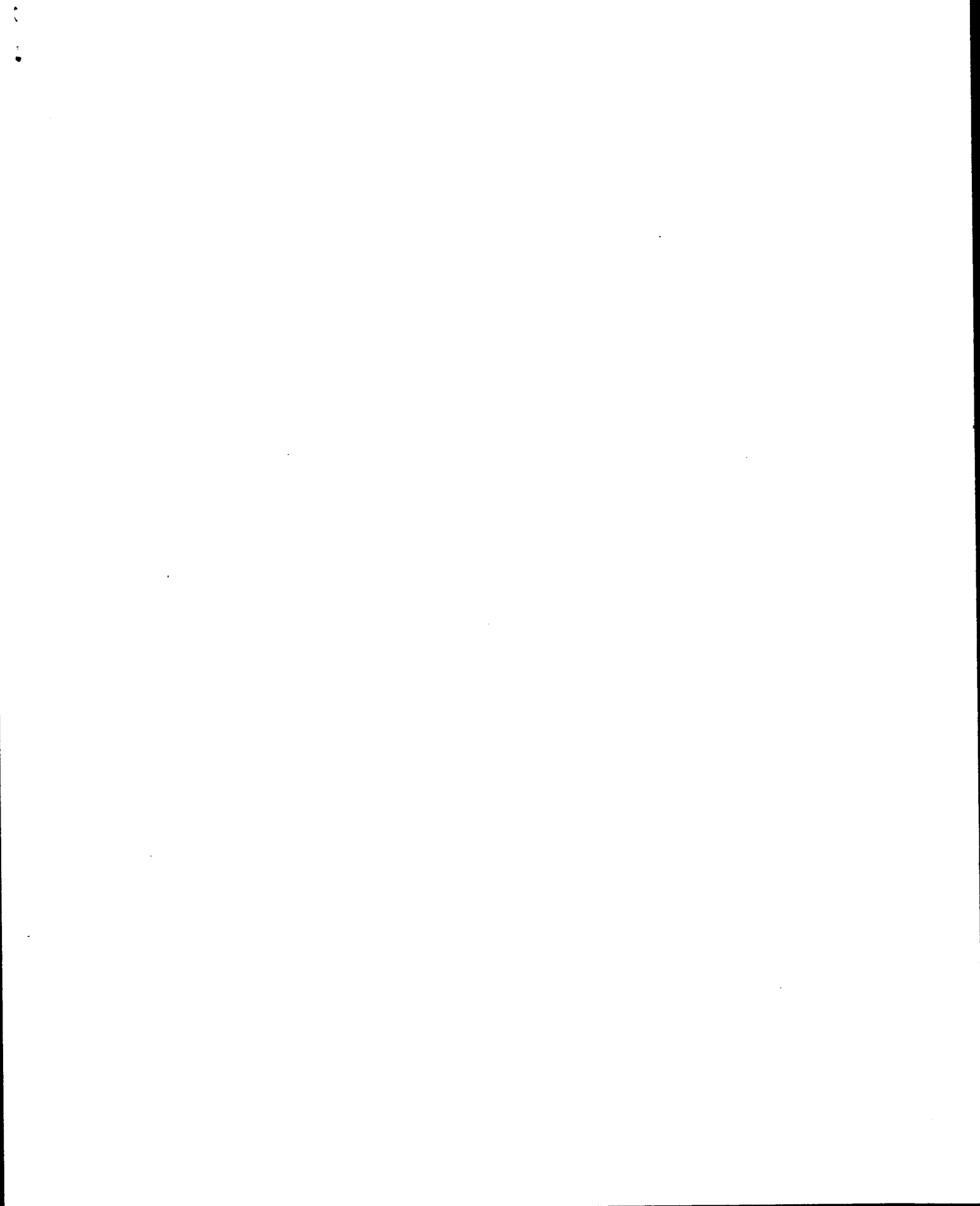
QY 226 gtcatgtgatgtgaagagagtgaaacaactgtctaccacccagagcagtgatctgg 285  
|||||  
Db 181 GTCATGTGATGTGTAAGGAGTGAACAACGTGGCTCTACCCGACGCACTGTCAATTTGG 240  
|||||

QY 286 tacccttggagaaatgcatgtccctttagtcttgacagtgltgcaaatccatcactcc 345  
|||||  
Db 241 TACCCTTGGAGAAAGCAATTCCTTAACTCTTGACAGTGTTCAAATTCATTCACCTCC 300  
|||||

QY 346 ttatttctgagaaactccctgtgttggagtgctccagtgaggaagagtgat 405  
|||||  
Db 301 TTATTTTCTGAGGAACCTCCTGTGTTTGGCAAGTTGGCTCCAGTGAGGAAGAGTGTAT 360  
|||||

QY 406 atgtgagggaaggaactcaactcagtggtttgaaagac 438  
|||||  
Db 361 ATGTAGGGAGGCAAACTCAGTGTGTAAGAC 393  
|||||

Search completed: August 18, 2001, 22:49:18  
Job time: 11880 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 18, 2001, 20:17:00 ; Search time 5437.46 Seconds  
(without alignments)  
2986.898 Million cell updates/sec

Title: US-09-284-320-31  
Perfect score: 1050  
Sequence: 1 atgctgtgtgtgtgtgtgt.....accagaagatcgaatgcat 1050

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 773874588 residues

Word size : 10

Total number of hits satisfying chosen parameters: 1131318

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_ba3:\*  
4: gb\_in1:\*  
5: gb\_in2:\*  
6: gb\_in3:\*  
7: gb\_in4:\*  
8: gb\_in5:\*  
9: gb\_in6:\*  
10: gb\_in7:\*  
11: gb\_in8:\*  
12: gb\_in9:\*  
13: gb\_in10:\*  
14: gb\_in11:\*  
15: gb\_in12:\*  
16: gb\_in13:\*  
17: gb\_in14:\*  
18: gb\_in15:\*  
19: gb\_in16:\*  
20: gb\_in17:\*  
21: gb\_in18:\*  
22: gb\_in19:\*  
23: gb\_in20:\*  
24: gb\_in21:\*  
25: gb\_in22:\*  
26: gb\_in23:\*  
27: gb\_in24:\*  
28: gb\_in25:\*  
29: gb\_in26:\*  
30: gb\_in27:\*  
31: gb\_in28:\*  
32: gb\_in29:\*  
33: gb\_in30:\*  
34: gb\_in31:\*  
35: gb\_in32:\*  
36: gb\_in33:\*  
37: gb\_in34:\*  
38: gb\_in35:\*  
39: gb\_in36:\*  
40: gb\_in37:\*  
41: gb\_in38:\*  
42: gb\_in39:\*  
43: gb\_in40:\*

44: em\_ov:\*  
45: em\_pat:\*  
46: em\_ph:\*  
47: em\_pl:\*  
48: em\_ro:\*  
49: em\_sts:\*  
50: em\_sy:\*  
51: em\_un:\*  
52: em\_v1:\*  
53: em\_v2:\*  
54: em\_v3:\*  
55: em\_v4:\*  
56: em\_v5:\*  
57: em\_v6:\*  
58: em\_v7:\*  
59: em\_v8:\*  
60: em\_v9:\*  
61: em\_v10:\*  
62: em\_v11:\*  
63: em\_v12:\*  
64: em\_v13:\*  
65: em\_v14:\*  
66: em\_v15:\*  
67: em\_v16:\*  
68: em\_v17:\*  
69: em\_v18:\*  
70: em\_v19:\*  
71: em\_v20:\*  
72: em\_v21:\*  
73: em\_v22:\*  
74: em\_v23:\*  
75: em\_v24:\*  
76: em\_v25:\*  
77: em\_v26:\*  
78: em\_v27:\*  
79: em\_v28:\*  
80: em\_v29:\*  
81: em\_v30:\*  
82: em\_v31:\*  
83: em\_v32:\*  
84: em\_v33:\*  
85: em\_v34:\*  
86: em\_v35:\*  
87: em\_v36:\*  
88: em\_v37:\*  
89: em\_v38:\*  
90: em\_v39:\*  
91: em\_v40:\*  
92: em\_v41:\*  
93: em\_v42:\*  
94: em\_v43:\*  
95: em\_v44:\*  
96: em\_v45:\*  
97: em\_v46:\*  
98: em\_v47:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	999	95.1	89	AF248966 Homo sapi
2	941	88.6	93	AL049929 Homo sapi
3	491	622	97	Y17975 Homo sapien
4	193	18.4	69	AC026156 Homo sapi
5	151	14.4	69	AC026156 Homo sapi
6	78	7.4	80	AL356315 Homo sapi
7	78	7.4	80	AL356381 Homo sapi
8	78	7.4	81	AL513325 Homo sapi

C 9	29	2.8	72831	94	AF259074	Mus muscu
C 10	22	2.1	17185	83	AP002424	Homo sapi
C 11	22	2.1	177097	82	AP001569	Homo sapi
C 12	22	2.1	179726	86	AC007052	Homo sapi
C 13	22	2.1	182411	78	AC009408	Homo sapi
C 14	22	2.1	196869	77	AC008753	Homo sapi
C 15	22	2.1	200774	82	AP001592	Homo sapi
C 16	22	2.1	326419	76	AC0079635	Homo sapi
C 17	21	2.0	90130	90	AL161618	Homo sapi
C 18	21	2.0	165699	82	AP001802	Homo sapi
C 19	21	2.0	177454	91	AP002853	Homo sapi
C 20	21	2.0	184653	62	AC012113	Homo sapi
C 21	21	2.0	185614	67	AC008104	Homo sapi
C 22	21	2.0	193279	87	AC009466	Homo sapi
C 23	21	2.0	227850	81	AL451143	Homo sapi
C 24	20	1.9	43410	5	CEC0106	
C 25	20	1.9	116448	12	AC005142	Arabidops
C 26	20	1.9	147620	12	AC001195	Homo sapi
C 27	20	1.9	149004	61	AC009445	Homo sapi
C 28	20	1.9	153097	81	AL512606	Homo sapi
C 29	20	1.9	159629	13	ATCRR1V9	AL161497 Arabidops
C 30	20	1.9	162064	90	AL1390882	Human DNA
C 31	20	1.9	163819	67	AC022758	Human DNA
C 32	20	1.9	168843	79	AL157772	Homo sapi
C 33	20	1.9	172601	66	AC021455	Homo sapi
C 34	20	1.9	172758	66	AC022553	Homo sapi
C 35	20	1.9	180977	66	AC020624	Homo sapi
C 36	20	1.9	185621	71	AC026624	Homo sapi
C 37	20	1.9	187982	82	AC036195	Homo sapi
C 38	20	1.9	190315	83	AL589844	Homo sapi
C 39	20	1.9	190802	66	CNS01D9	AL139022 Homo sapi
C 40	20	1.9	219888	70	AC007448	Homo sapi
C 41	20	1.9	223020	90	AC079959	Mus muscu
C 42	19	1.8	855	53	AL162424	Human DNA
C 43	19	1.8	870	96	OCMP	AL128614 clone BAO
C 44	19	1.8	898	94	AF201698	Y13551 Osetetlagia
C 45	19	1.8	1027	53	CNS07210	AF201698 Mus muscu
C 46	19	1.8	1439	4	AB025275	clone BAO
C 47	19	1.8	1660	15	SCYLR012C	AB025275 Sticliopu
C 48	19	1.8	1817	15	SCYLR013W	AB025275 Sticliopu
C 49	19	1.8	25818	63	AC014479	273185 S. cerevisia
C 50	19	1.8	75522	92	HS1168A5	AL031183 Human DNA
C 51	19	1.8	77378	64	AC016183	AL031683 Homo sapi
C 52	19	1.8	81207	92	HS12513	AL033828 Human DNA
C 53	19	1.8	104785	91	AP001465	AP001465 Homo sapi
C 54	19	1.8	108090	97	AC009533	AC009533 Homo sapi
C 55	19	1.8	110079	84	AC002315	AC002315 Mouse BAC
C 56	19	1.8	115400	81	AL391536	AL391536 Homo sapi
C 57	19	1.8	115402	82	AC012310	AP001967 Homo sapi
C 58	19	1.8	120407	83	AP001967	AP001967 Homo sapi
C 59	19	1.8	121703	86	AC006452	AC006452 Homo sapi
C 60	19	1.8	128769	87	AC013357	AC013357 Homo sapi
C 61	19	1.8	129467	93	HSJ7601K24	AL109626 Human DNA
C 62	19	1.8	137521	84	CNS06C82	AL1391156 Homo sapi
C 63	19	1.8	137945	78	AL133508	AL133508 Homo sapi
C 64	19	1.8	145772	61	AC010467	AC010467 Homo sapi
C 65	19	1.8	148485	64	AC016209	AC016209 Homo sapi
C 66	19	1.8	150785	69	AC026158	AC026158 Homo sapi
C 67	19	1.8	151770	93	HSAC002066	AC020266 Human BAC
C 68	19	1.8	152369	84	CNS07ED2	AL1454543 Homo sapi
C 69	19	1.8	153615	70	AC027168	AC027168 Homo sapi
C 70	1					





















```
Db 301 AGAACCCAGACAGTCCCTATACCTGATATAGTAAATTTGAAATATCCCTGCTT 360
QY 920 tcaactgttacttgatataatgccttgcccttgcttgatcattaccctctaca 979
Db 361 TCAACATGTGCTTGTGATATGATGCTTGGCTTGCTGCTGATTTATACCTTTACA 420
QY 980 atatttgacatgcatcctgatatgatacatcattataagatgacaacaagaaga 1039
Db 421 ATATTGGAAATGATGCTGATGATGATGATGATGATGATGATGATGATGATGAT 480
QY 1040 ttgaatgat 1050
Db 481 TTGCAATGAT 491

RESULT 4
AC026156 126543 bp DNA HTG 28-MAR-2000
LOCUS Homo sapiens chromosome 3p clone RP11-169016, WORKING DRAFT
DEFINITION SPOJENCE, 10 unordered pieces.
ACCESSION AC026156.2 GI:7331308
VERSION AC026156.2
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 126543)
AUTHORS Lu, G., Zhao, Y., Lin, W., Dong, H., Wan, M., Xu, S., Gu, W., Tu, Y.,
Jia, D., Wu, C., Zhang, C., Zhong, M., Zhou, Y., Ren, S., Fu, G., Chen, Z.
and Huang, M.
Direct Submission
Submitted (21-MAR-2000) Genomic Dept., Chinese National Human
Genome Center at Shanghai, 315 Guo Shou Jing Road, Shanghai,
Shanghai 201203, P. R. China
On Mar 28, 2000 this sequence version replaced gi:7271973.
-----Genome Center Information-----
Center: Chinese National Human Genome Center at Shanghai
Center Code: CHGC
Web site: http://www.chgc.sh.cn
Email: mhuan@chgc.sh.cn or fuan@chgc.sh.cn
-----End Genome Center Information-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 3411: contig of 3411 bp in length
* 3412 7987: contig of 4576 bp in length
* 7988 16490: contig of 8503 bp in length
* 16491 30218: contig of 13728 bp in length
* 30219 44895: contig of 14677 bp in length
* 44896 55628: contig of 10733 bp in length
* 55629 71833: contig of 16205 bp in length
* 71834 84537: contig of 12704 bp in length
* 84538 103706: contig of 19169 bp in length
* 103707 126543: contig of 22837 bp in length.
Location/Qualifiers
1. 126543
/organism="Homo sapiens"
/db_xref="taxon:9606"

Query Match 18.4%: Score 193; DB 69; Length 126543;
Best Local Similarity 100.0%; Pred. NO. 2.9e-101;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 858 gaagaccacgaagaagtcctcctataaccccttgacatataaglaatttgaatctcgtgt 917
Db 81178 GAAGAACCCAGACAGTCCCTATACCTGATATAGTAAATTTGAAATATCCCTGCT 81237
QY 918 ttcaactgttacttgatataatgccttgcccttgcttgatcattaccctctaca 977
Db 81238 TTCAACATGTGCTTGTGATATGATGCTTGGCTTGCTGCTGATTTATACCTTTACA 81297
QY 978 caatatttgacatgcatcctgatatgatacatcattataagatgacaacaagaaga 1037
Db 81298 CAATATTGGAAATGATGCTGATGATGATGATGATGATGATGATGATGATGATGAT 81357
QY 1038 gattcgaaatgat 1050
Db 81358 GATTGCAATGAT 81370

RESULT 5
AC026156 126543 bp DNA HTG 28-MAR-2000
LOCUS Homo sapiens chromosome 3p clone RP11-169016, WORKING DRAFT
DEFINITION SPOJENCE, 10 unordered pieces.
ACCESSION AC026156.2 GI:7331308
VERSION AC026156.2
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 126543)
AUTHORS Lu, G., Zhao, Y., Lin, W., Dong, H., Wan, M., Xu, S., Gu, W., Tu, Y.,
Jia, D., Wu, C., Zhang, C., Zhong, M., Zhou, Y., Ren, S., Fu, G., Chen, Z.
and Huang, M.
Direct Submission
Submitted (21-MAR-2000) Genomic Dept., Chinese National Human
Genome Center at Shanghai, 315 Guo Shou Jing Road, Shanghai,
Shanghai 201203, P. R. China
On Mar 28, 2000 this sequence version replaced gi:7271973.
-----Genome Center Information-----
Center: Chinese National Human Genome Center at Shanghai
Center Code: CHGC
Web site: http://www.chgc.sh.cn
Email: mhuan@chgc.sh.cn or fuan@chgc.sh.cn
-----End Genome Center Information-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 3411: contig of 3411 bp in length
* 3412 7987: contig of 4576 bp in length
* 7988 16490: contig of 8503 bp in length
* 16491 30218: contig of 13728 bp in length
* 30219 44895: contig of 14677 bp in length
* 44896 55628: contig of 10733 bp in length
```



SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
AUTHORS 1 (bases 1 to 165617)  
TITLE Plumb.B.  
JOURNAL Direct Submission  
Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On Sep 9, 2000 this sequence version replaced gi:9797513.  
----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquerry@sanger.ac.uk  
Project Information  
Center project name: BA294A4  
----- Summary Statistics  
Assembly program: XGAP4: version 4.5  
Sequencing vector: plasmid: L08752: 100% of reads  
Chemistry: Dye-terminator big dye: 100% of reads  
Consensus quality: 159696 bases at least Q40  
Consensus quality: 161504 bases at least Q30  
Consensus quality: 162803 bases at least Q20  
Insert size: 164017: sum-of-contigs  
Insert size: 174497: 3.2% error: agarose-fp  
Quality coverage: 4.29x in Q20 bases: sum-of-contigs Quality  
coverage: 4.21x in Q20 bases: agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 17 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1  
8110: contig of 8110 bp in length  
8111 8210: gap of 100 bp  
8211 13854: contig of 5644 bp in length  
13855 13954: gap of 100 bp  
13955 46993: contig of 33039 bp in length  
46994 47093: gap of 100 bp  
47094 59514: contig of 12421 bp in length  
59515 59614: gap of 100 bp  
59615 86817: contig of 27203 bp in length  
86818 86917: gap of 100 bp  
86918 90065: contig of 3148 bp in length  
90066 90165: gap of 100 bp  
90166 101858: contig of 11693 bp in length  
101859 101958: gap of 100 bp  
101959 105470: contig of 3512 bp in length  
105471 105570: gap of 100 bp  
105571 109111: contig of 3541 bp in length  
109112 109211: gap of 100 bp  
109212 113589: contig of 4378 bp in length  
113590 113689: gap of 100 bp  
113690 124809: contig of 11120 bp in length  
124810 124909: gap of 100 bp  
124910 128323: contig of 3414 bp in length  
128324 128423: gap of 100 bp  
128424 140194: contig of 11771 bp in length  
140195 140294: gap of 100 bp  
140295 149122: contig of 8828 bp in length  
149123 149222: gap of 100 bp  
149223 154697: contig of 5475 bp in length  
154698 154797: gap of 100 bp  
154798 157559: contig of 2762 bp in length  
157560 157659: gap of 100 bp  
157660 165617: contig of 7958 bp in length.  
Location/Qualifiers  
1. 165617  
FEATURES  
source

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP11-294A4"  
/clone\_lib="RP11-11.2"  
1. 8110  
/note="assembly\_fragment:01010  
fragment\_chain:1"  
8211. 13854  
/note="assembly\_fragment:00790  
fragment\_chain:1"  
13955. 46993  
/note="assembly\_fragment:00399  
fragment\_chain:1"  
47094. 59514  
/note="assembly\_fragment:00426  
fragment\_chain:1"  
59615. 86817  
/note="assembly\_fragment:00065  
fragment\_chain:1"  
86918. 90065  
/note="assembly\_fragment:00014  
fragment\_chain:1"  
90166. 101858  
/note="assembly\_fragment:00559  
fragment\_chain:1"  
101959. 105470  
/note="assembly\_fragment:01342  
fragment\_chain:2"  
105571. 109111  
/note="assembly\_fragment:01371  
fragment\_chain:2"  
109212. 113589  
/note="assembly\_fragment:01100  
fragment\_chain:2"  
113690. 124809  
/note="assembly\_fragment:00361  
fragment\_chain:2"  
124910. 128323  
/note="assembly\_fragment:00588  
fragment\_chain:3"  
128424. 140194  
/note="assembly\_fragment:00969  
fragment\_chain:3"  
140295. 149122  
/note="assembly\_fragment:00693  
fragment\_chain:3"  
149223. 154697  
/note="assembly\_fragment:01864  
fragment\_chain:4"  
154798. 157559  
/note="assembly\_fragment:01778  
fragment\_chain:4"  
157660. 165617  
/note="assembly\_fragment:00012"  
/note="assembly\_fragment:00012"  
BASE COUNT 52833 a 30460 c 29131 g 51583 t 1610 others  
ORIGIN

Query Match 74%; Score 78; DB 80; Length 165617;  
Best local similarity 100.0%; Pred. No. 8.2e-34; Indels 0; Gaps 0;  
Matches 78; Conservative 0; Mismatches 0;  
QY 596 gtcataagcattcagcagaagtcattctcattatattatattatctgagctgagcaggtt 655  
|||||  
Db 133324 GTCATAAACATCTAGCAAGATCTCTCTGATTATATTCAGCTGAGCTGCGAGGTT 133265  
QY 656 tggatgaattggagc 673  
|||||  
Db 133264 TGGATGAATTGGAGAC 133247  
RESULT 8

```

AL513325/LOCUS      AL513325      214692 bp      DNA      HTG      09-MAR-2001
DEFINITION      Homo sapiens chromosome 1 clone RP11-332L8, *** SEQUENCING IN
ACCESSION      AL513325
VERSION      AL513325.5 GI:13274893
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 214692)
McLay, K.
Direct Submission
Submitted (03-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
Requests: clonerequests@sanger.ac.uk
On Mar 12, 2001 this sequence version replaced gi:1325041.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BA332L8
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 209611 bases at least Q40
Consensus quality: 21375 bases at least Q30
Consensus quality: 21265 bases at least Q20
Insert size: 213792; sum-of-ctigs
Quality coverage: 6.07x in Q20 bases; sum-of-ctigs Quality
coverage: 6.75x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 13451: contig of 13451 bp in length
* 13452 13551: gap of 100 bp
* 13552 48859: contig of 35308 bp in length
* 48860 48959: gap of 100 bp
* 48960 51722: contig of 2763 bp in length
* 51723 51822: gap of 100 bp
* 51823 66224: contig of 14402 bp in length
* 66225 66324: gap of 100 bp
* 66325 85895: contig of 19571 bp in length
* 85896 85995: gap of 100 bp
* 85996 117920: contig of 31925 bp in length
* 117921 118020: gap of 100 bp
* 118021 168860: contig of 50840 bp in length
* 168861 168960: gap of 100 bp
* 168961 177592: contig of 8632 bp in length
* 177593 177692: gap of 100 bp
* 177693 193973: contig of 16281 bp in length
* 193974 194073: gap of 100 bp
* 194074 214692: contig of 20619 bp in length.
Location/Qualifiers
1. 214692
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-332L8"
/clone_lib="RPC1-11.2"
1. 13451
/note="assembly_fragment:03382

```

---

```

misc_feature      fragment_chain:1"
13552..48859
/note="assembly_fragment:03000
fragment_chain:1"
misc_feature      48960..51722
/note="assembly_fragment:02751"
51823..66224
/note="assembly_fragment:00639
fragment_chain:2"
66325..85895
/note="assembly_fragment:02639
fragment_chain:2"
85996..117920
/note="assembly_fragment:03009
fragment_chain:2"
118021..168860
/note="assembly_fragment:01179
fragment_chain:2"
168961..177592
/note="assembly_fragment:00474
fragment_chain:2"
177693..193973
/note="assembly_fragment:00396
fragment_chain:2"
194074..214692
/note="assembly_fragment:01410
fragment_chain:2
clone_end:SP6
vector_side:right"
BASE COUNT      69445 a 38568 c 38843 g 66931 t 905 others
ORIGIN
Query Match      7.4%; Score 78; DB 81; Length 214692;
Best Local Similarity 100.0%; Pred. No. 8.2e-34;
Matches 78; Conservative 0; Mismatches 0; Indels 0;
QY 596 gtcaataagctcgaagatcctcctgattatattcactgagctgcaggtt 655
|||||
Db 199311 GTCATAGCATCTAGCCAGGATCATTCTCGATTATATTCACGAGCGACGCTT 199252
|||||
QY 656 tggatgaattggagac 673
|||||
Db 199251 TGCATCAATTGGAGAC 199234
|||||
RESULT 9
AF259074/c LOCUS      AF259074/c
DEFINITION      14D1-D2, complete sequence.
ACCESSION      AF259074
VERSION      AF259074.1 GI:8575574
KEYWORDS      house mouse.
SOURCE      Mus musculus
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 72831)
REFERENCE      Lee,I.Y., Wang,K., Lasky,S.R., Dahl,T., Hall,J. and Hood,L.E.
AUTHORS      Large-scale Sequence Analysis of the Mouse T-cell Receptor
TITLE      Alpha/Delta Locus
JOURNAL      Unpublished
2 (bases 1 to 72831)
REFERENCE      Lee,I.Y., Wang,K., Lasky,S.R., Dahl,T., Hall,J. and Hood,L.E.
AUTHORS      Direct Submission
TITLE      Submitted (21-APR-2000) Department of Molecular Biotechnology, Box
JOURNAL      35730 University of Washington, Seattle, WA 98195, USA
COMMENT      Interspersed Repeats were identified with RepeatMasker (available
from http://ftp.genome.washington.edu/RM/RepeatMasker.html) Simple
sequence repeats were identified with sputnik (available from
http://serac.mbt.washington.edu/chrisa/software/sputnik.html).
Location/Qualifiers

```

```

source
1. 72831
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="14"
/map="14D1-D2"
/clone="MBAC519; RG:298L19"
/cell_line="CJ7"
/clone_id="Research Genetics Mouse BAC library"
/ote="T-cell receptor alpha locus; flanking sequences are
UM:MBAC1058 (GenBank Accession Number AF259072) 5' and
UM:MBAC01 (GenBank Accession Number AF259071) 3'."
complement(1..3813)
repeat_region
/rpt_family="L1_MM"
complement(3794..9677)
repeat_region
/rpt_family="L1_MM"
complement(9585..10193)
repeat_region
/rpt_family="L1F_Send"
complement(9922..10405)
repeat_region
/rpt_family="L1F_Send"
complement(10134..10617)
repeat_region
/rpt_family="L1F_Send"
complement(10346..10666)
repeat_region
/rpt_family="L1F_Send"
complement(10688..11092)
repeat_region
/rpt_family="L1M2_of12"
complement(11099..11689)
repeat_region
/rpt_family="L1_RN"
11945..11972
repeat_region
/rpt_family="CA)n"
12294..12490
repeat_region
/rpt_family="RLTR13D"
13228..13278
repeat_region
/rpt_family="TGG)n"
13291..13464
repeat_region
/rpt_family="TGG)n"
14374..15028
repeat_region
/rpt_family="RLTR13C"
14515..15028
repeat_region
/rpt_family="RLTR13D"
15157..15862
repeat_region
/rpt_family="L1_MM"
15873..16356
repeat_region
/rpt_family="L1F_Send"
16085..16568
repeat_region
/rpt_family="L1F_Send"
16509..16992
repeat_region
/rpt_family="L1F_Send"
16933..17541
repeat_region
/rpt_family="L1F_Send"
17449..23342
repeat_region
/rpt_family="L1_MM"
23348..23928
repeat_region
/rpt_family="L1_MM"
23365..24582
repeat_region
/rpt_family="L1_RN"
24584..24701
repeat_region
/rpt_family="L1_MM"
24599..26220
repeat_region
/rpt_family="L1_MM"
complement(26217..27218)
repeat_region
/rpt_family="L1_MM"
complement(27118..27357)
repeat_region
/rpt_family="L1F_Send"
27881..28092
repeat_region
/rpt_family="B3"
complement(28543..28587)
repeat_region
/rpt_family="CA)n"
complement(28929..28969)
repeat_region
/rpt_family="CA)n"
28938..28969
repeat_region
/rpt_unit="(GT)x15
29384..29531
repeat_region
/rpt_family="WTD"
repeat_region
29556..29866
/rpt_family="L1_RN"
29875..30280
repeat_region
/rpt_family="RLTR13C"
30282..31643
repeat_region
/rpt_family="L1_MM"
31674..32181
repeat_region
/rpt_family="RLTR13C"
32186..34760
repeat_region
/rpt_family="L1_MM"
34747..36147
repeat_region
/rpt_family="L1_RN"
34929..34953
repeat_region
/rpt_unit="(AAC)x6
34930..34953
repeat_region
/rpt_family="CAAA)n"
36156..36306
repeat_region
/rpt_family="Lx3"
36337..36557
repeat_region
/rpt_family="RMER13A"
36561..37025
repeat_region
/rpt_family="RMER13A"
37029..37841
repeat_region
/rpt_family="Lx4"
37821..38430
repeat_region
/rpt_family="Lx4"
complement(38431..38476)
repeat_region
/rpt_family="AT-rich"
39113..39169
repeat_region
/rpt_unit="(CA)x28
39114..39169
repeat_region
/rpt_family="CA)n"
complement(39309..40536)
repeat_region
/rpt_family="L1_MM"
complement(40537..40629)
repeat_region
/rpt_family="Lx5"
40636..40687
repeat_region
/rpt_family="RMER6A"
40688..41215
repeat_region
/rpt_family="RMER6A"
complement(41325..41441)
repeat_region
/rpt_family="RMER1B"
41452..41573
repeat_region
/rpt_family="ORR1A3"
42854..42964
repeat_region
/rpt_family="BI-F"
42986..43037
repeat_region
/rpt_family="CAAA)n"
complement(44550..44693)
repeat_region
/rpt_family="RMER1B"
complement(44699..44893)
repeat_region
/rpt_family="RMER1B"
44957..45152
repeat_region
/rpt_family="MERVL"
45066..45164
repeat_region
/rpt_family="HERVL"
45202..46018
repeat_region
/rpt_family="MERVL"
complement(46029..46760)
repeat_region
/rpt_family="L1_MM"
46762..46985
repeat_region
/rpt_family="MERVL"
complement(46986..47171)
repeat_region
/rpt_family="ORR1A1"
complement(47015..47353)
repeat_region
/rpt_family="ORR1C"
47201..47258
repeat_region
/rpt_unit="(AG)x28
47202..47258
repeat_region
/rpt_family="CA)n"
complement(47350..47483)
repeat_region
/rpt_family="RMER1B"
complement(47643..48403)
repeat_region

```

```

repeat_region /rpt_family="RMR1B"
complement(48459..48558)
repeat_region /rpt_family="ORR1C"
48573..48751
repeat_region /rpt_family="MERVL"
48611..48933
repeat_region /rpt_family="HERVL"
48940..49259
repeat_region /rpt_family="L1_MM"
49288..49549
repeat_region /rpt_family="MERVL"
complement(49570..49703)
repeat_region /rpt_family="(TR)n"
49751..49830
repeat_region /rpt_family="B1-F"
complement(49884..50083)
repeat_region /rpt_family="LIME2"
complement(49885..50110)
repeat_region /rpt_family="LIME1"
51858..51878
repeat_region /rpt_family="AT-rich"
52211..52307
repeat_region /rpt_family="PBI9"
52224..52294
repeat_region /rpt_family="B1_MM"
52319..52422
repeat_region /rpt_family="LIME2_orf2"
52424..52487
repeat_region /rpt_family="B1-F"

```

## Query Match

Best Local Similarity 2.88; Score 29; DB 94; Length 72831;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 325 gttgcaattccattcattcatttc 353
Db 43798 gttgcaattccattcattcatttc 43770
|||||

```

```

RESULT 10
AP002424/c 171185 bp DNA HTG 31-MAY-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-873122 map 18q21, WORKING
DRAFT SEQUENCE, 18 unordered pieces.
ACCESSION AP002424.1 GI:8131688
VERSION HTG: HTGS.PHASE1; HTGS.DRAFT.
KEYWORDS Homo sapiens DNA, clone:RP11-873122.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 171185)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 171,185 genomic DNA of 18q21
Published Only in Database (2000) In press
2 (bases 1 to 171185)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (29-MAY-2000) Masahito Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsc.riken.go.jp, Tel:81-42-778-9923,
URL:http://hgp.gsc.riken.go.jp/, Fax:81-42-778-9924)
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDrafl18

```

Center clone name: RP11-873122  
----- Summary Statistics  
Sequencing vector: PCR products; 100% of reads  
Assembly: Dye-terminator ET-amersham; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 15697 bases at least Q40  
Consensus quality: 164110 bases at least Q30  
Consensus quality: 167421 bases at least Q20  
Insert size: 169485; sum-of-ctrls  
Quality coverage: 4.80x in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently consists of 18 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1 24923 contig of 24923 bp in length
25024 47004 contig of 21981 bp in length
47105 67070 contig of 19966 bp in length
67171 82757 contig of 15987 bp in length
82858 95544 contig of 12687 bp in length
95645 109592 contig of 13948 bp in length
109693 120167 contig of 10475 bp in length
120168 129255 contig of 8988 bp in length
129256 129355 contig of 100 bp in length
129356 137003 contig of 7648 bp in length
137004 144376 contig of 7273 bp in length
144377 144476 contig of 100 bp in length
144477 150659 contig of 6183 bp in length
150660 150759 contig of 100 bp in length
150760 155896 contig of 5137 bp in length
155897 155996 contig of 100 bp in length
155997 160522 contig of 4526 bp in length
160523 160622 contig of 100 bp in length
160623 163263 contig of 2641 bp in length
163264 163363 contig of 100 bp in length
163364 165194 contig of 1831 bp in length

```

NOTE: This is a 'working draft' sequence. It currently consists of 18 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

FEATURES

- \* 165195 165294: gap of 100 bp
- \* 165295 167704: contig of 2410 bp in length
- \* 167705 167804: gap of 100 bp
- \* 167805 168750: contig of 1946 bp in length
- \* 169751 169850: gap of 100 bp
- \* 169851 171185: contig of 1335 bp in length.

Location/Qualifiers

1..171185

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="18"

/map="18q21"

/clone="RP11-873L22"

1..24923

misc\_feature /note="assembly\_fragment" 25024..47004

misc\_feature /note="assembly\_fragment" 47105..67070

misc\_feature /note="assembly\_fragment" 67171..82757

misc\_feature /note="assembly\_fragment" 82838..95544

misc\_feature /note="assembly\_fragment" 95645..109592

misc\_feature /note="assembly\_fragment" 109693..120167

misc\_feature /note="assembly\_fragment" 120268..129255

misc\_feature /note="assembly\_fragment" 129356..137003

misc\_feature /note="assembly\_fragment clone\_end:SP6 vector\_side:right" 137104..144376

misc\_feature /note="assembly\_fragment" 144477..150659

misc\_feature /note="assembly\_fragment" 150760..155896

misc\_feature /note="assembly\_fragment" 155997..160522

misc\_feature /note="assembly\_fragment" 160623..163263

misc\_feature /note="assembly\_fragment" 163364..165194

misc\_feature /note="assembly\_fragment clone\_end:T7 vector\_side:left" 165295..167704

misc\_feature /note="assembly\_fragment" 167805..169750

misc\_feature /note="assembly\_fragment" 169851..171185

misc\_feature /note="assembly\_fragment"

BASE COUNT 52031 a 32086 c 33289 g 52078 t 1701 others

ORIGIN

Query Match 2.1% Score 22 DB 83 Length 171185

Best Local Similarity 100.0% Pred. No. 0.58 Indels 0 Gaps 0

Matches 22: Conservative 0 Mismatches 0

Oy 889 tataagataattgataatt 910

|||||

Db 15070 TATAAGTATATTTGAAATTT 15049

RESULT 11

AP001569/c AP001569 177097 bp DNA HTG 30-MAY-2000

LOCUS Homo sapiens chromosome 18 clone RP11-859C21 map 18q21. WORKING

DEFINITION DRAFT SEQUENCE, 32 unordered pieces.

ACCESSION AP001569

VERSION AP001569.2 GI:8117403

KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.

SOURCE Homo sapiens DNA, clone:RP11-859C21.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

1 (bases 1 to 177097)

Author(s) Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujii,Yama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

Journal Homo sapiens 177,097 genomic DNA of 18q21

Published Only in Database (2000) In press

2 (bases 1 to 177097)

Author(s) Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujii,Yama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

Direct Submission

Submitted (29-MAR-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan (E-mail:hattori@gsr.riken.go.jp, URL:http://hqp.gsc.riken.go.jp/, Tel:81-42-778-9923, Fax:81-42-778-9924)

On May 30, 2000 this sequence version replaced gi:7380904.

## COMMENT

----- Genome Center

Center: RIKEN Genomic Sciences Center(GSC)

Center code: RIKEN

Web site: http://hqp.gsc.riken.go.jp/

Contact: hattori@gsr.riken.go.jp

----- Project Information

Center project name: RP11-859C21

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads

Chemistry: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 160670 bases at least Q40

Consensus quality: 168058 bases at least Q30

Consensus quality: 171631 bases at least Q20

Insert size: 173997; sum-of-contigs

Quality coverage: 4.63x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 32 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

23893	23792	contig of	23792	bp in length
45465	contig of	21573	bp in length	
58167	contig of	12602	bp in length	
58268	contig of	12906	bp in length	
71274	contig of	9980	bp in length	
81354	contig of	9727	bp in length	
91181	contig of	6207	bp in length	
97488	contig of	7415	bp in length	
105003	contig of	6105	bp in length	
111208	contig of	6644	bp in length	
116952	contig of	6634	bp in length	
123686	contig of	5300	bp in length	
129086	contig of	5148	bp in length	
134334	contig of	4255	bp in length	
138689	contig of	3885	bp in length	
142674	contig of	3613	bp in length	
146387	contig of	3657	bp in length	
150053	contig of	3893	bp in length	
154014	contig of	2079	bp in length	
156326	contig of	2253	bp in length	
158578	contig of	2446	bp in length	
161124	contig of	2335	bp in length	
161225	contig of	1879	bp in length	
163660	contig of	1311	bp in length	
165639	contig of	1191	bp in length	
168240	contig of	1233	bp in length	
168341	contig of	1299	bp in length	
169680	contig of	1002	bp in length	
171079	contig of	1334	bp in length	
172181	contig of	1000	bp in length	
173615	contig of	1262	bp in length	
174715	contig of	1262	bp in length	





```

Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C.,
Furie,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heatford,A., Horton,L., Howland,J.C., Jones,C., Kamm,L.,
Kartas,A., Lehoczy,J., Lieu,C., Locke,K., MacDonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Strange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Tastaye,S., Toriella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Direct Submission
Submitted (09-MAR-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 179726)

Bairren,B., Linton,L., Nusbau,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckery,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Depayre,E., Fitzhugh,W., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., Fitchugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heatford,A., Horton,L., Howland,J.C., Jones,C., Kamm,L.,
Kartas,A., Lehoczy,J., Lieu,C., Locke,K., MacDonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Tastaye,S., Toriella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.

Direct Submission
Submitted (23-MAR-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 24, 1999 this sequence version replaced gi:443872.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.

LOCATION/Qualifiers
1. 179726
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RPIC-11 human BAC library"
/map="18"
/chromosome="18"
898..1014
/rpl_family="L1MD3"
complement(1035..1333)
/rpl_family="AluSx"
2987..3030
/rpl_family="Cr-rich"
3087..3291
/rpl_family="L1MC4"
4298..4330
/rpl_family="(CA)n"
5344..5365
/rpl_family="AT-rich"
complement(5552..5724)
/rpl_family="AluJ"
5731..11847
/rpl_family="L1PA2"
11857..11891
/rpl_family="(TAA)n"
complement(11893..12002)
/rpl_family="AluJ"
11925..11948
/Note="Single-stranded coverage."
repeat_region
12026..12321
/rpl_family="AluSg"
complement(12605..12846)
/rpl_family="MIR"

repeat_region
13352..13660
/rpl_family="AluSc"
complement(13741..14128)
/rpl_family="MLT1B"
14425..14456
/rpl_family="MLT1B"
14457..14978
/rpl_family="AT-rich"
15496..15834
/rpl_family="MLT1F"
16325..16628
/rpl_family="MLT1B"
16753..17086
/rpl_family="AluY98"
17543..17601
/rpl_family="MER7A"
18075..18442
/rpl_family="MIR"
18075..18442
/rpl_family="MLT1B"
complement(18735..18882)
/rpl_family="MIR"
complement(18948..19083)
/rpl_family="L1MB8"
complement(19190..19426)
/rpl_family="AluJo"
19431..19750
/rpl_family="L1MB3"
20434..20555
/rpl_family="MLT1A1"
20556..20863
/rpl_family="AluSx"
20864..21080
/rpl_family="MLT1A1"
21528..21565
/rpl_family="AT-rich"
21575..21668
/rpl_family="L2"
complement(21849..21991)
/rpl_family="MER3B"
23147..23168
/rpl_family="(TTTG)n"
complement(23169..23384)
/rpl_family="AluJo"
23385..23419
/rpl_family="AT-rich"
complement(23799..23846)
/rpl_family="L2"
complement(23947..24060)
/rpl_family="L2"
24066..24206
/rpl_family="BC200"
complement(24347..24514)
/rpl_family="MER21B"
24782..25203
/rpl_family="L1MB2"
25385..26005
/rpl_family="L1MB2"
26035..26336
/rpl_family="AluSx"
26432..26604
/rpl_family="L1MB"
26610..26645
/rpl_family="AT-rich"
complement(27113..27422)
/rpl_family="AluSp"
27865..28190
/rpl_family="AluSg"
28250..28272
/rpl_family="AT-rich"
28363..28496
/rpl_family="HL1"
complement(28497..28796)
/rpl_family="AluSg"
28797..28869

```

repeat_region	/rpt_family="HALL1"
repeat_region	28993. .29027
repeat_region	/rpt_family="AT_rich"
repeat_region	29107. .29134
repeat_region	/rpt_family="TANATGn"
repeat_region	29151. .29182
repeat_region	/rpt_family="(G)n"
repeat_region	complement(29821. .30164
repeat_region	/rpt_family="MER1A"
repeat_region	complement(30236. .30522
repeat_region	/rpt_family="AluDo"
repeat_region	30576. .30723
repeat_region	/rpt_family="MR"
repeat_region	30927. 31271
repeat_region	/rpt_family="MITTK"
repeat_region	32362. .32437
repeat_region	/rpt_family="(CA)n"
repeat_region	32480. .32502
repeat_region	/rpt_family="(TTG)n"
repeat_region	complement(32503. .32751)
repeat_region	/rpt_family="AluSc"
repeat_region	34312. .34476
repeat_region	/rpt_family="MR"
repeat_region	complement(34648. .34953)
repeat_region	/rpt_family="AluSc"
repeat_region	complement(35550. .35947)
repeat_region	/rpt_family="HEHB"
repeat_region	36048. .36070
repeat_region	/rpt_family="(TTTA)n"
repeat_region	complement(36072. .36366)
repeat_region	/rpt_family="AluDo"
repeat_region	36525. .36587
repeat_region	/rpt_family="AT_rich"
repeat_region	36792. .36849
repeat_region	/rpt_family="AT_rich"
repeat_region	complement(36903. .37092)
repeat_region	/rpt_family="LMA7"
repeat_region	37502. .37812

Query Match	2.18;	Score 22;	DB 86;	Length 179726;
Best Local Similarity	100.0%;	Pred. No. 0.58;		
Matches 22;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

```
QY      889  cataagtataatttgaatatt 910
          |||||
Db 28319  TATAGTATAATTTCGAATATT 28298
```

RESULT	13	
AC090408/c		
LOCUS		
DEFINITION	AC090408 182411 bp DNA	HTG 16-MAR-2001
	Homo sapiens chromosome 18 clone RP11-859C21	map 18, WORKING DRAFT
ACCSSION	AC090408	SRUNCHE, 14 unordered pieces.
VERSION	AC090408.2	GI:13357356
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.	
SOURCE	human.	

REFERENCE	Mammalia; Eutheria; Primates; Carnivora; Felidae; Felinae; Felis concolor.
AUTHORS	1 (bases 1 to 18,411)
TITLE	Homologous recombination between nonhomologous chromosomes in the domestic cat ( <i>Felis catus</i> )
JOURNAL	Genetics
REFERENCE	2 (bases 1 to 18,411)
AUTHORS	Birren, B., Linton, L., Nussbaum, C., Lander, E., Allien, N., Anderson S., Birren, B., Linton, L., Nussbaum, C., Lander, E., Allien, N., Anderson S.

Birren, B., Linton, L., Nusbach, C., Lander, E., Allen, N., Anderson, S.,  
 Barina, N., Bastien, V., Boguslavsky, L., Boukhalter, B., Brown, A.,  
 Camarata, C., Campopiano, A., Choepoy, Y., Colangelo, M., Collins, S.,  
 Conradi, A., Cooke, P., DeArallano, K., Dewar, K., Diaz, J. S.,  
 Dodd, S., Fazio, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,  
 Garayza, S., Gagne, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Harcus, B., Hefford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
 Jones, C., Karats, A., Latouche, K., Lameiras, R., Landers, T.,

TITLE  
JOURNAL  
COMMENT

Leloczky, J., Levine, R., Liu, G., Maclean, C., Macdonald, P.,  
 Marquis, N., Matthews, R., McCarthy, M., McEwen, P., McKernan, K.,  
 McPeckers, R., Meldrum, J., Menzies, I., Mihova, T., Mienna, V.,  
 Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,  
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
 Phunhkanh, P., Pierre, N., Pollara, V., Raymond, C., Reiter, R.,  
 Riach, M., Riley, R., Rise, C., Rogov, P., Roman, J., Roselli, M.,  
 Roy, A., Santos, R., Schauer, S., Schumpack, R., Seaman, S., Severy, P.,  
 Soungez, C., Spencer, B., Stange-Thumann, N., Stojanovic, N.,  
 Strauss, N., Subramaniam, A., Talamas, J., Testeyle, H., Theodore, J.,  
 Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A.,  
 Wilson, B., Wu, X., Wyman, D., Ye, M., Young, G., Zainoun, J.,  
 Zembek, L., Zimmer, A. and Zody, M.J.  
 Direct Submission  
 Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Mar 16, 2001 this sequence version replaced gi:12958044.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/MW/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: W1BR  
 Web site: <http://www.seq.wi.mit.edu>  
 Contacts: [sequence-submissions@genome.wi.mit.edu](mailto:sequence-submissions@genome.wi.mit.edu)  
 Project Information  
 Center project name: L12718  
 Center clone name: 859.C.21  
 Summary Statistics

Sequencing vector	Plasmid; N/A; 100% of reads
Chemistry	Dye-terminator Big Dye; 100% of reads
Assembly program	Phrap; 0.960731
Consensus quality	177348 bases at least Q40
Consensus quality	179859 bases at least Q30
Consensus quality	180657 bases at least Q20
Insert size	18111; sum-of-configs
Quality coverage	6.2 in Q20 bases; sum-of-conf

NOTE: This is a "working draft" sequence. It currently consists of 14 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

*	1	803:	contig of 803 bp in length
*	804	903:	gap of 100 bp
*	904	1787:	contig of 884 bp in length
*	1788	1887:	gap of 100 bp
*	1888	3003:	contig of 1116 bp in length
*	3004	3103:	gap of 100 bp
*	3104	5488:	contig of 2393 bp in length
*	5489	5596:	gap of 100 bp
*	5599	7587:	contig of 1969 bp in length
*	7588	7687:	gap of 100 bp
*	7688	9759:	contig of 2052 bp in length
*	9740	9839:	gap of 100 bp
*	9840	13254:	contig of 3415 bp in length
*	13255	13354:	gap of 100 bp
*	13355	17739:	contig of 4385 bp in length
*	17740	17839:	gap of 100 bp
*	17840	25230:	contig of 7391 bp in length
*	25231	25330:	gap of 100 bp
*	25331	32077:	contig of 6747 bp in length
*	32078	32177:	gap of 100 bp
*	32178	48716:	contig of 16539 bp in length
*	48717	48816:	gap of 100 bp
*	48817	68998:	contig of 20182 bp in length
*	68989	69098:	gap of 100 bp
*	69099	103010:	contig of 33912 bp in length
*	103011	103110:	gap of 100 bp
*	103111	182411:	contig of 79301 bp in length



FEATURES \* 151220 196869: contig of 45650 bp in length.  
Location/Qualifiers  
source 1..196869  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="18"  
/map="18"  
/clone="RP11-869L2"  
/clone\_lib="RPC1-11 Human Male BAC"  
misc\_feature 1..11489  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left  
misc\_feature 11590..12389  
/note="assembly\_fragment"  
misc\_feature 12490..15301  
/note="assembly\_fragment"  
misc\_feature 15402..17297  
/note="assembly\_fragment"  
misc\_feature 17398..22220  
/note="assembly\_fragment"  
misc\_feature 22321..27991  
/note="assembly\_fragment"  
misc\_feature 28092..31242  
/note="assembly\_fragment"  
misc\_feature 31343..41247  
/note="assembly\_fragment"  
misc\_feature 41348..49982  
/note="assembly\_fragment"  
misc\_feature 50083..64395  
/note="assembly\_fragment"  
misc\_feature 64496..79369  
/note="assembly\_fragment"  
misc\_feature 79470..151119  
/note="assembly\_fragment"  
misc\_feature 151220..196869  
/note="assembly\_fragment"  
clone\_end:77  
vector\_side:right

BASE COUNT 56352 a 38146 c 39166 g 61991 t 1214 others

ORIGIN

Query Match 2.1%: Score 22; DB 77; Length 196869;  
Best Local Similarity 100.0%: Pred. No. 0.58;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 889 tataagataattgaatt 910  
|||||  
Db 122196 TATAGTATATTTGAATAT 122217

RESULT 15  
AP001592  
LOCUS AP001592 200774 bp DNA HTG 29-JUL-2000  
DEFINITION Homo sapiens chromosome 18 clone RP11-869L2 map 18q21, WORKING  
DRAFT SEQUENCE, 11 unordered pieces.  
ACCESSION AP001592  
VERSION AP001592.3 GI:9588073  
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.  
SOURCE Homo sapiens DNA, clone:RP11-869L2.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 200774)  
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
TITLE Homo sapiens 200,774 genomic DNA of 18q21  
JOURNAL Published Only in Database (2000) In press  
2 (bases 1 to 200774)  
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
TITLE Direct Submission

## JOURNAL

Submitted (31-MAR-2000) Masahira Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan  
(E-mail:hattori@gsc.riken.go.jp, URL:htp://hgp.gsc.riken.go.jp/,  
Tel:81-42-778-9923, Fax:81-42-778-9924)  
On Jul 31, 2000 this sequence version replaced g1:8117416.  
----- Genome Center  
Center: RIKEN Genomic Sciences Center (GSC)  
Center code: RIKEN  
Web site: htp://hgp.gsc.riken.go.jp/  
Contact: hattori@gsc.riken.go.jp  
----- Project Information  
Center Project name: HumDrat118  
Center Clone name: RP11-869L2  
----- Summary Statistics  
Sequencing vector: PCR products: 100% of reads  
Chemistry: Dye-terminator ET-amersham: 100% of reads  
Assembly program: Phrap: version 0.990329  
Consensus quality: 197367 bases at least Q40  
Consensus quality: 196735 bases at least Q30  
Consensus quality: 199244 bases at least Q20  
Insert size: 199774; sum-of-contigs  
Quality coverage: 9.59x in Q20 bases; sum-of-contigs

## COMMENT

NOTE: This is a 'working draft' sequence. It currently consists of  
11 contigs. The true order of the pieces is not known and their  
order in this sequence record is arbitrary. Gaps between the  
contigs are represented as runs N, but the exact sizes of the gaps  
are unknown. This record will be updated with the finished sequence  
as soon as it is available and the accession number will be  
preserved.

1 36889 contig of 36889 bp in length  
36990 73347 contig of 36358 bp in length  
73448 98904 contig of 25457 bp in length  
99005 124595 contig of 25591 bp in length  
124696 145130 contig of 20435 bp in length  
145231 163732 contig of 18502 bp in length  
163833 181811 contig of 17979 bp in length  
181912 190895 contig of 8984 bp in length  
190996 198298 contig of 7303 bp in length  
198399 199648 contig of 1250 bp in length  
199749 200774 contig of 1026 bp in length.  
NOTE: This is a 'working draft' sequence. It currently  
consists of 11 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 36889: contig of 36889 bp in length  
\* 36890 36989: gap of 100 bp  
\* 36990 73347: contig of 36358 bp in length  
\* 73448 73447: gap of 100 bp  
\* 73448 98904: contig of 25457 bp in length  
\* 98905 99004: gap of 100 bp  
\* 99005 124595: contig of 25591 bp in length  
\* 124596 124695: gap of 100 bp  
\* 124696 145130: contig of 20435 bp in length  
\* 145131 145230: gap of 100 bp  
\* 145231 163732: contig of 18502 bp in length  
\* 163733 163832: gap of 100 bp  
\* 163833 181811: contig of 17979 bp in length  
\* 181812 181911: gap of 100 bp  
\* 181912 190895: contig of 8984 bp in length  
\* 190896 190995: gap of 100 bp  
\* 190996 198298: contig of 7303 bp in length  
\* 198299 198398: gap of 100 bp  
\* 198399 199648: contig of 1250 bp in length  
\* 199649 199748: gap of 100 bp  
\* 199749 200774: contig of 1026 bp in length.  
Location/Qualifiers  
1..200774

FEATURES  
source

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q21"
/clone="RP11-869L2"
1. .36889
/misc_feature
/note="assembly-fragment clone_end:T7 vector_side:left"
36990. .73347
/misc_feature
/note="assembly-fragment"
73448. .98904
/misc_feature
/note="assembly-fragment"
99005. .124595
/misc_feature
/note="assembly-fragment"
124696. .145130
/misc_feature
/note="assembly-fragment"
145231. .163732
/misc_feature
/note="assembly-fragment"
163833. .181811
/misc_feature
/note="assembly-fragment"
181912. .190895
/misc_feature
/note="assembly-fragment"
190996. .198298
/misc_feature
/note="assembly-fragment"
198399. .199648
/misc_feature
/note="assembly-fragment clone_end:SP6 vector_side:left"
199749. .200774
BASE COUNT 61549 a 39812 c 38979 g 59434 t 1000 others
ORIGIN
Query Match 2.1%; Score 22; DB 82; Length 200774;
Best Local Similarity 100.0%; Pred. No. 0.58;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 889 tataaataattgaattt 910
Db 97681 TATAGTATATTTGATTT 97702

RESULT 16
AC079635 326419 bp DNA HTG 06-SEP-2000
LOCUS AC079635/c
DEFINITION Mus musculus clone Rp23-152L20, WORKING DRAFT SEQUENCE, 80
unordered pieces.
ACCESSION AC079635
VERSION AC079635.1 GI:9972296
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 326419)
McCombie,W.R., Baker,J.P., Bahret,A., Bal,H., Dedhia,N.N., de la
Bastide,M., Huang,E.N., King,L., Kirchoff,K.A., Miller,B.,
Naschimoto,L.O., O'Shaughnessy,A.L., Preston,R.R., Rodriguez,M.A.,
Shah,R.S., Shekher,M., Spiegel,L.A., Toth,K. and Vill,M.D.
Mouse Genomic Sequence
Unpublished
2 (bases 1 to 326419)
McCombie,W.R.
Direct Submission
Submitted (06-SEP-2000) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
----- Genome Center
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
Laboratory
Center code: CSHL
Web site: http://www.cshl.org/genseq
Contact: mccombie@cschl.org
----- Project Information
Center project name: Rp23-152L20
```

```
Center clone name: Rp23-152L20
*****
* NOTE: This is a 'working draft' sequence. It currently
* consists of 80 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 26314: contig of 26314 bp in length
* 26315 26386: gap of unknown length
* 26387 39292: contig of 12906 bp in length
* 39293 39364: gap of unknown length
* 39365 50228: contig of 10864 bp in length
* 50229 50300: gap of unknown length
* 50301 60933: contig of 10633 bp in length
* 60934 61005: gap of unknown length
* 61006 70882: contig of 9877 bp in length
* 70883 70954: gap of unknown length
* 70955 79815: contig of 8861 bp in length
* 79816 79887: gap of unknown length
* 79888 88738: contig of 8851 bp in length
* 88739 88810: gap of unknown length
* 88811 97340: contig of 8530 bp in length
* 97341 97412: gap of unknown length
* 97413 105541: contig of 8129 bp in length
* 105542 113643: gap of unknown length
* 113644 113715: contig of 8030 bp in length
* 113716 121706: contig of 7991 bp in length
* 121707 121778: gap of unknown length
* 121779 128971: contig of 7193 bp in length
* 128972 129043: gap of unknown length
* 129044 134621: contig of 5578 bp in length
* 134622 134693: gap of unknown length
* 134694 140231: contig of 5538 bp in length
* 140232 140303: gap of unknown length
* 140304 145558: contig of 5255 bp in length
* 145559 145630: gap of unknown length
* 145631 150800: contig of 5170 bp in length
* 150801 150872: gap of unknown length
* 150873 155900: contig of 5118 bp in length
* 155901 156062: gap of unknown length
* 156063 160780: contig of 4718 bp in length
* 160781 160852: gap of unknown length
* 160853 165352: contig of 4500 bp in length
* 165353 165424: gap of unknown length
* 165425 169815: contig of 4391 bp in length
* 169816 169887: gap of unknown length
* 169888 174083: contig of 4196 bp in length
* 174084 174155: gap of unknown length
* 174156 178123: contig of 3967 bp in length
* 178124 178194: gap of unknown length
* 178195 181919: contig of 3725 bp in length
* 181920 181991: gap of unknown length
* 181992 185620: contig of 3629 bp in length
* 185621 185692: gap of unknown length
* 185693 189321: contig of 3629 bp in length
* 189322 189393: gap of unknown length
* 189394 192915: contig of 3522 bp in length
* 192916 192987: gap of unknown length
* 192988 196498: contig of 3511 bp in length
* 196499 200064: gap of unknown length
* 200065 200136: gap of unknown length
* 200137 203624: contig of 3488 bp in length
* 203625 203696: gap of unknown length
* 203697 207120: contig of 3424 bp in length
* 207121 207192: gap of unknown length
* 207193 210614: contig of 3422 bp in length
* 210615 210686: gap of unknown length
* 210687 213955: contig of 3269 bp in length
```

```

* 213956 214027: gap of unknown length
* 214028 217239: contig of 3212 bp in length
* 217240 217311: gap of unknown length
* 217312 220495: contig of 3184 bp in length
* 220496 220567: gap of unknown length
* 220568 223735: contig of 3168 bp in length
* 223736 223806: gap of unknown length
* 223807 226972: contig of 3166 bp in length
* 226973 227043: gap of unknown length
* 227044 230116: contig of 3073 bp in length
* 230117 233252: contig of 3065 bp in length
* 233253 233323: gap of unknown length
* 233324 236212: contig of 2889 bp in length
* 236213 236283: gap of unknown length
* 236284 239039: contig of 2756 bp in length
* 239040 239110: gap of unknown length
* 239111 241796: contig of 2686 bp in length
* 241797 241867: gap of unknown length
* 241868 244536: contig of 2669 bp in length
* 244537 244607: gap of unknown length
* 244608 247255: contig of 2648 bp in length
* 247256 247326: gap of unknown length
* 247327 249839: contig of 2513 bp in length
* 249840 249910: gap of unknown length
* 249911 252421: contig of 2511 bp in length
* 252422 252492: gap of unknown length
* 252493 254997: contig of 2505 bp in length
* 254998 255068: gap of unknown length
* 255069 257570: contig of 2502 bp in length
* 257571 257641: gap of unknown length
* 257642 260103: contig of 2462 bp in length
* 260104 260174: gap of unknown length
* 260175 262347: contig of 2373 bp in length
* 262348 262618: gap of unknown length
* 262619 264990: contig of 2372 bp in length
* 264991 265061: gap of unknown length
* 265062 267415: contig of 2354 bp in length
* 267416 267486: gap of unknown length
* 267487 269770: contig of 2284 bp in length
* 269771 269841: gap of unknown length
* 269842 272123: contig of 2282 bp in length
* 272124 272194: gap of unknown length
* 272195 274472: contig of 2278 bp in length
* 274473 274543: gap of unknown length
* 274544 276808: contig of 2265 bp in length
* 276809 276879: gap of unknown length
* 276880 279134: contig of 2255 bp in length
* 279135 279205: gap of unknown length
* 279206 281445: contig of 2240 bp in length
* 281446 281516: gap of unknown length
* 281517 283749: contig of 2233 bp in length
* 283750 283820: gap of unknown length
* 283821 285992: contig of 2172 bp in length
* 285993 286063: gap of unknown length
* 286064 288234: contig of 2171 bp in length
* 288235 288305: gap of unknown length
* 288306 290465: contig of 2161 bp in length
* 290467 290537: gap of unknown length
* 290538 292685: contig of 2148 bp in length
* 292686 292756: gap of unknown length
* 292757 294903: contig of 2147 bp in length
* 294904 294974: gap of unknown length
* 294975 297149: contig of 2104 bp in length
* 297150 299243: gap of unknown length
* 299244 299314: contig of 2094 bp in length
* 299315 301404: gap of unknown length
* 301405 301475: gap of unknown length
* 301476 303530: contig of 2055 bp in length
* 303531 305626: gap of unknown length
* 305627 305697: gap of unknown length

```

```

Query Match
Best Local Similarity 100.0%; Score 22; DB 76; Length 326419;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 391 gaggaagagtgatatgtag 412
Db 252266 GAGGAAGAAGTATATGTAG 252245
|||||

```

```

RESULT 17
AL161618/c
LOCUS
DEFINITION
Human DNA sequence from clone RP11-268F1 on chromosome 6 Contains
ESTs, STSS, GSSs and a Cpg island. Contains the MEPIA gene for
meprin A alpha (PABA peptide hydrolase), complete sequence.
ACCESSION
AL161618
VERSION
AL161618.17 GI:10443401
KEYWORDS
HTG: Cpg island; MEPIA; meprin; peptide hydrolase.
SOURCE
human
ORANISM
Homo sapiens

```

```

REFERENCE
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 90130)
AUTHORS
Ramsey/H.
TITLE
Submitted (09-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
requests: clonerequest@sanger.ac.uk
JOURNAL
CB10 ISA, UK. E-mail enquiries: humquerry@sanger.ac.uk

```

## COMMENT

On Oct 1, 2000 this sequence version replaced gi:10119713. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>

IMPORTANT: This sequence is not the entire insert of clone RP11-268F1. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP11-268F1 is at 1 in this sequence. The true left end of clone RP3-365012 is at 90031 in this sequence.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. RP11-268F1 is from the library RP11-1 constructed by the group of Pletier de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6.

FEATURES	Location/qualifiers
1	00130

**source**

repeat_region	/organism="Homo sapiens"	
repeat_region	/db_xref="taxon:9606"	
repeat_region	/chromosome="6"	
repeat_region	/clone="RP11-268P1"	
repeat_region	/clone.lib="RPC1-11.1"	
repeat_region	404. .646	
repeat_region	/note="MIR repeat: matches 6. .256 of consensus"	
repeat_region	1603. .1672	
repeat_region	/note="MIR repeat: matches 76. .145 of consensus"	
repeat_region	complement(1904. .2544)	
repeat_region	/note="match: GSS: Em:AQ350405"	
repeat_region	2494. .2876	
repeat_region	/note="match: GSS: Em:AQ341945"	
repeat_region	3235. .3410	
repeat_region	/note="L1M2 repeat: matches 5986. .6163 of consensus"	
repeat_region	3896. .4311	
repeat_region	/note="L2 repeat: matches 1198. .1641 of consensus"	
repeat_region	5090. .5321	
repeat_region	/note="L2 repeat: matches 2485. .2704 of consensus"	
repeat_region	6624. .6933	
repeat_region	/note="Charlie5 repeat: matches 2257. .2585 of consensus"	
repeat_region	complement(7279. .7766)	
repeat_region	/note="match: GSS: Em:AQ838062"	
repeat_region	7506. .7545	
repeat_region	/note="10 copies 4 mer taga 100% conserved"	
repeat_region	7507. .7546	
repeat_region	/note="20 copies 2 mer ag 77% conserved"	
repeat_region	7735. .7869	
repeat_region	/note="MIR repeat: matches 123. .261 of consensus"	
repeat_region	8223. .8353	
repeat_region	/note="MIR repeat: matches 85. .224 of consensus"	
repeat_region	8380. .8586	
repeat_region	/note="MIR repeat: matches 41. .256 of consensus"	
repeat_region	9125. .10036	
repeat_region	/note="L1P4 repeat: matches 5233. .6145 of consensus"	
repeat_region	10344. .10411	
repeat_region	/note="L2 repeat: matches 2668. .2736 of consensus"	
repeat_region	10563. .10936	
repeat_region	/note="MT1B repeat: matches 1. .359 of consensus"	
repeat_region	11792. .12109	
repeat_region	/note="AluX8 repeat: matches 1. .309 of consensus"	
repeat_region	12385. .12505	
repeat_region	/note="MIR repeat: matches 21. .141 of consensus"	
repeat_region	12520. .12622	
repeat_region	/note="L2 repeat: matches 2364. .2483 of consensus"	
repeat_region	13106. .13537	
repeat_region	/note="match: GSS: Em:B62917"	
repeat_region	13227. .13533	
repeat_region	/note="AluX8 repeat: matches 1. .307 of consensus"	
repeat_region	14167. .16741	
repeat_region	/note="L1M1 repeat: matches 3476. .6154 of consensus"	
repeat_region	16310. .16791	
repeat_region	/note="match: GSS: Em:B79103"	
repeat_region	16904. .17074	
repeat_region	/note="MER5B repeat: matches 1. .178 of consensus"	
repeat_region	17122. .17590	
repeat_region	/note="Charlie4 repeat: matches 25. .491 of consensus"	
repeat_region	17773. .18083	
repeat_region	/note="AluX8 repeat: matches 1. .302 of consensus"	
repeat_region	18286. .18345	
repeat_region	/note="Alu repeat: matches 1. .48 of consensus"	
repeat_region	18331. .18382	

repeat_region	/note="Alu repeat: matches 254. .303 of consensus" 18383. .18404 /note="11 copies 2 mer aa 100% conserved"
repeat_region	18669. .18702 /note="17 copies 2 mer ac 82% conserved"
repeat_region	18951. .19113 /note="MIR repeat: matches 2. .187 of consensus"
repeat_region	19506. .19603 /note="MIR repeat: matches 73. .173 of consensus"
repeat_region	19689. .19804 /note="MIR repeat: matches 33. .147 of consensus"
misc_feature	/note="match: GSS: Em:AQ331606" 20795. .21041
misc_feature	/note="L1M7 repeat: matches 5916. .6171 of consensus" complement(20840. .21312) /note="match: GSS: Em:AQ37608"
misc_feature	complement(20948. .21297) /note="match: GSS: Em:AQ38394"
repeat_region	21082. .21775 /note="12 repeat: matches 2010. .2710 of consensus"
repeat_region	21766. .22653 /note="PTRS repeat: matches 768. .1626 of consensus"
misc_feature	21939. .22375 /note="CpG island" /evidence=not_experimental
repeat_region	22664. .23296 /note="LIR12 repeat: matches 1. .671 of consensus"
repeat_region	23327. .23514 /note="MIR repeat: matches 50. .247 of consensus"
repeat_region	23670. .23759 /note="Charlie4 repeat: matches 1855. .1945 of consensus"
repeat_region	23889. .24060 /note="12 repeat: matches 2548. .2746 of consensus"
repeat_region	24065. .24481 /note="MIR2 repeat: matches 1. .388 of consensus"
repeat_region	24482. .25021 /note="12 repeat: matches 1967. .2534 of consensus"
repeat_region	25395. .25934 /note="L1M82 repeat: matches 5590. .6155 of consensus"
repeat_region	26223. .26664 /note="L1PB2 repeat: matches 5278. .5736 of consensus"
repeat_region	26680. .27027 /note="114 copies 2 mer at 82% conserved"
repeat_region	26685. .27020 /note="24 copies 14 mer 83% conserved"
repeat_region	26687. .27022 /note="12 copies 28 mer 83% conserved"
repeat_region	27029. .27218 /note="L1PB2 repeat: matches 5784. .5968 of consensus"
repeat_region	27219. .27514 /note="Alus repeat: matches 1. .298 of consensus"
misc_feature	27515. .27666 /note="L1PB2 repeat: matches 5968. .6154 of consensus" complement(27834. .28277)
repeat_region	/note="match: GSS: Em:AQ465424" 29379. .29573
repeat_region	/note="65 copies 3 meraat 59% conserved" 29399. .29566 /note="12 copies 14 mer 64% conserved"
repeat_region	29420. .29503 /note="21 copies 4 mer ataa 64% conserved"
repeat_region	29536. .29894 /note="12 repeat: matches 1816. .2093 of consensus"
repeat_region	30835. .30973 /note="MIR1A2 repeat: matches 238. .374 of consensus"
repeat_region	30974. .31254 /note="Alus repeat: matches 1. .295 of consensus"
repeat_region	31261. .31430 /note="MIR1A2 repeat: matches 1. .171 of consensus"
misc_feature	33214. .33622 /note="match: GSS: Em:B50816"
repeat_region	33678. .34033

```

repeat_region /note="LIMEI repeat: matches 5791. .6163 of consensus"
34164. .34297
/note="1/3 repeat: matches 2577. .2710 of consensus"
repeat_region 34431. .34567
/note="MIR repeat: matches 119. .261 of consensus"
repeat_region 36071. .36109
/note="MERSA repeat: matches 53. .91 of consensus"
misc_feature 36117. .36505
/note="match: GSS: Em:A0661796"
repeat_region 36303. .36489
/note="MERSA repeat: matches 1. .184 of consensus"
repeat_region 36663. .36837
/note="MIR repeat: matches 81. .256 of consensus"
repeat_region 36911. .37117
/note="MERSA repeat: matches 1. .224 of consensus"
repeat_region 37913. .38026

```

```

Query Match 2.0%; Score 21; DB 90; Length 90130;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 229 atgctgattggaaggaggt 249
|||||
Db 28131 ATGCTGATTGGAAGGAGTG 28111

```

```

RESULT 18
AP001802 165699 bp DNA HMG 08-AUG-2000
LOCUS Homo sapiens chromosome 11 clone RP11-100J10 map 11q, WORKING DRAFT
DEFINITION
SEQUENCE 11 unordered pieces.
ACCESSION AP001802
VERSION AP001802.3 GI:9757489
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens DNA, clone:RP11-100J10.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 165699)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Tokoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 165,699 genomic DNA of 11q
Published Only in Database (2000) In press
2 (bases 1 to 165699)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Tokoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (14-APR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(e-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
tel:81-42-778-9923, fax:81-42-778-9924)
On Aug 9, 2000 this sequence version replaced gi:8117473.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center Project name: RP11-100J10
Center clone name: RP11-100J10
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 162162 bases at least Q40
Consensus quality: 163565 bases at least Q40
Consensus quality: 164180 bases at least Q20
Insert size: 164699; sum-of-ctrls
Quality coverage: 11.49x in Q20 bases; sum-of-ctrls

```

NOTE: This is a 'working draft' sequence. It currently consists of 11 contigs. The true order of the pieces is not known and their

order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1 39447 contig of 39447 bp in length
39548 71742 contig of 32195 bp in length
71843 99207 contig of 27365 bp in length
99308 117932 contig of 18625 bp in length
118033 134393 contig of 16361 bp in length
134494 139889 contig of 5396 bp in length
139990 145501 contig of 5512 bp in length
145602 150056 contig of 4455 bp in length
150157 155960 contig of 4455 bp in length
156061 161995 contig of 5804 bp in length
162096 165699 contig of 5935 bp in length
162096 165699 contig of 3604 bp in length.

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 39447: contig of 39447 bp in length
39448 39547: gap of 100 bp
39548 71742: contig of 32195 bp in length
71743 71842: gap of 100 bp
71843 99207: contig of 27365 bp in length
99208 99307: gap of 100 bp
99308 117932: contig of 18625 bp in length
117933 118032: gap of 100 bp
118033 134393: contig of 16361 bp in length
134394 134493: gap of 100 bp
134494 139889: contig of 5396 bp in length
139890 139989: gap of 100 bp
139990 145501: contig of 5512 bp in length
145502 145601: gap of 100 bp
145602 150056: contig of 4455 bp in length
150057 150156: gap of 100 bp
150157 155960: contig of 5804 bp in length
155961 156060: gap of 100 bp
156061 161995: contig of 5935 bp in length
161996 162095: gap of 100 bp
162096 165699: contig of 3604 bp in length.

```

#### FEATURES

```

Source
1. 165699
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-100J10"
1. 39447
/note="assembly-fragment"
39548. 71742
/note="assembly-fragment"
71843. 99207
/note="assembly-fragment"
99308. 117932
/note="assembly-fragment"
118033. 134393
/note="assembly-fragment"
134494. 139889
/note="assembly-fragment clone_end:SP6 vector_side:right"
139990. 145501
/note="assembly-fragment clone_end:SP7 vector_side:left"
145602. 150056
/note="assembly-fragment"
150157. 155960
/note="assembly-fragment"
156061. 161995
/note="assembly-fragment"
162096. 165699

```



BASE COUNT 44435 a 35614 c 36787 g 47863 t 1000 others  
 ORIGIN

Query Match 2.0% Score 21; DB 82; Length 165699;  
 Best Local Similarity 100.0%; Pred. No. 2.2;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 969 cacccttaacaatttgaa 989  
 Db 130150 CACCTCTTACAAATTTGGAA 130170

RESULT 19  
 AP002853/C  
 LOCUS  
 DEFINITION Homo sapiens genomic DNA, chromosome 11q clone:RP11-46K15, complete  
 sequences.

ACCESSION AP002853 GI:13094219  
 VERSION HTG  
 KEYWORDS Homo sapiens DNA, clone:RP11-46K15.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens

REFERENCE 1 (sites)  
 AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
 Fujiyama,A., Yada,T., Totoki,Y., Matanabe,H. and Sakaki,Y.  
 TITLE Homo sapiens genomic DNA  
 JOURNAL Published only in Database (2000) In press  
 REFERENCE 2 (bases 1 to 177464)  
 AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
 Fujiyama,A., Yada,T., Totoki,Y., Matanabe,H. and Sakaki,Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (15-AUG-2000) Masahira Hattori, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC),  
 1-7-22 Saitoh-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 (E-mail:hattori@gsc.riken.go.jp, URL:http://ngp.gsc.riken.go.jp/,  
 Tel:81-45-503-9111, Fax:81-45-503-9170)  
 On Feb 21, 2001 this sequence version replaced g1:1136467.

COMMENT  
 FEATURES  
 source  
 1. 177464  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="11"  
 /map="11q"  
 /clone="RP11-46K15"  
 BASE COUNT 50963 a 35528 c 36516 g 52457 t  
 ORIGIN

Query Match 2.0% Score 21; DB 91; Length 177464;  
 Best Local Similarity 100.0%; Pred. No. 2.2;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 969 cacccttaacaatttgaa 989  
 Db 149519 CACCTCTTACAAATTTGGAA 149499

RESULT 20  
 AC012113  
 LOCUS  
 DEFINITION Homo sapiens chromosome 11 clone RP11-195C14 map 11, WORKING DRAFT  
 SEQUENCE, 5 unordered pieces.  
 ACCESSION AC012113  
 VERSION AC012113.3 GI:7523735  
 KEYWORDS HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 184663)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 TITLE Homo sapiens chromosome 11, clone RP11-195C14  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 184663)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
 Baldino,J., Barne,N., Beckert,R., Boguslavsky,L., Boukigalter,B.,  
 Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,  
 Cooke,P., DeCarliano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,  
 Ferreira,P., Fitzhugh,M., Forrest,C., Funke,R., Gage,D., Horton,L.,  
 Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
 Lehoczy,J., Liu,C., Locke,K., Macdonald,P., Marguis,N.,  
 McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,  
 Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
 Strange-Thomson,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
 Wyman,D., Ye,M.J., Zimmer,A. and Zody,M.

TITLE  
 JOURNAL  
 COMMENT  
 Submitted (20-OCT-1999) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 7, 2000 this sequence version replaced g1:6649509.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 Project Information  
 Center project name: 1372  
 Center clone name: 195.C.14

Summary Statistics  
 Sequencing vector: M13; M77815; 100% of reads  
 Chemistry: Dye-Primer-amersham; 2% of reads  
 Chemistry: Dye-terminator Big Dye; 98% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 182733 bases at least Q40  
 Consensus quality: 183581 bases at least Q30  
 Consensus quality: 183925 bases at least Q20  
 Insert size: 170000; agarose-fp  
 Insert size: 184263; sum-of-contigs  
 Quality coverage: 10.1 in Q20 bases; agarose-fp  
 Quality coverage: 9.3 in Q20 ba.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 5 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 1099: contig of 1099 bp in length  
 \* 1100 1199: gap of 100 bp  
 \* 1200 14453: contig of 13254 bp in length  
 \* 14454 14553: gap of 100 bp  
 \* 14554 31589: contig of 17036 bp in length  
 \* 31590 31689: gap of 100 bp  
 \* 31690 68796: contig of 37107 bp in length  
 \* 68797 68896: gap of 100 bp  
 \* 68897 184663: contig of 115767 bp in length.

FEATURES  
 source  
 1. 184663  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="11"  
 /map="11"  
 /clone="RP11-195C14"  
 /clone-lib="RPCT-11 Human Male BAC"  
 1. 1099  
 /note="assembly\_fragment"

```

misc_feature      1200..14453
                  /note="assembly_fragment"
misc_feature      14554..31589
                  /note="assembly_fragment"
                  clone_end:SP6
misc_feature      31690..68796
                  /note="assembly_fragment"
                  /note="assembly_fragment"
                  clone_end:T7
misc_feature      68897..184663
                  /note="assembly_fragment"
                  vector_side:left"
BASE COUNT      54473 a 39976 c 37123 g 52691 t      400 others
ORIGIN

```

```

Query Match      2.0%; Score 21; DB 62; Length 184663;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 969 cacccttcacatttgga 989
Db 103423 CACCTTTCACATTTGGA 103443
|||||

```

```

RESULT 21
AC008104/c      185614 bp      DNA      HTG      26-MAY-2000
LOCUS
DEFINITION      Homo sapiens chromosome 18 clone RP11-446F17 map 18, WORKING DRAFT
ACCESSION      AC008104
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
TITLE      Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL      Homo sapiens chromosome 18, clone RP11-446F17
REFERENCE
AUTHORS      2 (bases 1 to 185614)
Unpublished

```

```

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barua,N., Beckert,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearliano,K., Depayre,E., Devon,K., Dewar,K.,
Doelelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lebecky,J., Liu,C., Locke,K., MacDonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Nifliff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Tasfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (22-JUL-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced gi:7321473.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: W1BR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1022
Center clone name: 446.F.17
----- Summary Statistics

```

```

Sequencing vector: M13; M77815; 92% of reads
Sequencing vector: Plasmid; n/a: %0.1% of reads
7.56620428751576Chemistry: Dye-primer-amesham; 92% of reads
Chemistry: Dye-terminator Big Dye; 8% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 166796 bases at least Q40
Consensus quality: 174175 bases at least Q30
Consensus quality: 177996 bases at least Q20
Insert size: 192000; agarose-1p
Insert size: 183514; sum-of-contigs
Quality covera.
NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1109: contig of 1109 bp in length
1110 1209: gap of 100 bp
1210 3242: contig of 2033 bp in length
3243 3342: gap of 100 bp
3343 4984: contig of 1642 bp in length
4985 5084: gap of 100 bp
5085 8149: contig of 3065 bp in length
8150 8249: gap of 100 bp
8250 10485: contig of 2236 bp in length
10486 10585: gap of 100 bp
10586 12579: contig of 1994 bp in length
12580 12679: gap of 100 bp
12680 15104: contig of 2425 bp in length
15105 15204: gap of 100 bp
15205 19613: contig of 4409 bp in length
19614 19713: gap of 100 bp
19714 23297: contig of 3584 bp in length
23298 23397: gap of 100 bp
23398 27445: contig of 4048 bp in length
27446 27545: gap of 100 bp
27546 33569: contig of 6024 bp in length
33570 33669: gap of 100 bp
33670 39886: contig of 6217 bp in length
39887 39986: gap of 100 bp
39987 46130: contig of 6144 bp in length
46131 46230: gap of 100 bp
46231 53403: contig of 7173 bp in length
53404 53503: gap of 100 bp
53504 61709: contig of 8206 bp in length
61710 61809: gap of 100 bp
61810 70390: contig of 8581 bp in length
70391 70490: gap of 100 bp
70491 83671: contig of 13181 bp in length
83672 83771: gap of 100 bp
83772 99035: contig of 15264 bp in length
99036 99135: gap of 100 bp
99136 114877: contig of 15742 bp in length
114878 114977: gap of 100 bp
114978 132866: contig of 17889 bp in length
132867 132966: gap of 100 bp
132967 157662: contig of 24696 bp in length
157663 157762: gap of 100 bp
157763 185614: contig of 27852 bp in length.
Location/Qualifiers
1. 185614
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-446F17"
/clone_lib="RP11-11 Human Male BAC"
1. 1109
/note="assembly_fragment"
1210. 3242

```

```

misc_feature
misc_feature
misc_feature

```

```

misc_feature /note="assembly_fragment"
3343..4984
/note="assembly_fragment"
misc_feature 5085..8149
/note="assembly_fragment"
misc_feature 8250..10485
/note="assembly_fragment"
misc_feature 10586..12579
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
misc_feature 12680..15104
/note="assembly_fragment"
misc_feature 15205..19613
/note="assembly_fragment"
misc_feature 19714..23297
/note="assembly_fragment"
misc_feature 23398..27445
/note="assembly_fragment"
misc_feature 27546..33569
/note="assembly_fragment"
misc_feature 33670..39886
/note="assembly_fragment"
misc_feature 39987..46130
/note="assembly_fragment"
clone_end:17
vector_side:right"
46231..53403
/note="assembly_fragment"
misc_feature 53504..61709
/note="assembly_fragment"
misc_feature 61810..70390
/note="assembly_fragment"
misc_feature 70491..83671
/note="assembly_fragment"
misc_feature 83772..99035
/note="assembly_fragment"
misc_feature 99136..114877
/note="assembly_fragment"
misc_feature 114978..132866
/note="assembly_fragment"
misc_feature 132967..157662
/note="assembly_fragment"
misc_feature 157763..185614
/note="assembly_fragment"
BASE COUNT 56535 a 55534 c 35072 g 56364 t 2109 others
ORIGIN

```

```

Query Match 2.0%; Score 21; DB 60; Length 185614;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 229 atggtgagtgatgaaggagtg 249
|||||
DB 99452 ATGCGTATGCTGAAGGAGTG 99432

```

```

RESULT 22
AC009466/c AC009466 193279 bp DNA PRI 15-DEC-2000
LOCUS Homo sapiens chromosome 11, clone RP11-87N22, complete sequence.
ACCESSION AC009466
VERSION AC009466.17 GI:11321807
KEYWORDS RTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 193279)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 11, clone RP11-87N22
JOURNAL Unpublished

```

```

REFERENCE
AUTHORS 2 (bases 1 to 193279)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckwith,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearrellano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C.,
Funk,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagob,B., Heatford,A., Horton,L., Locke,K., Macdonald,P.,
Karatas,A., Lehoczy,J., Liu,C., McKernan,K., McLaughlin,J.,
Marquis,N., McEwan,P., McGuire,A., Morrow,J., Mychalecky,J.,
Meldrim,J., Molla,M., Morris,W., Morrow,P., Pavlin,B.,
Naylor,J., Nijhoff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Stefyke,S., Toruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
REFERENCE
AUTHORS 3 (bases 1 to 193279)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A.,
Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearrellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Hiley,I., Johnson,R.,
Hagos,B., Heatford,A., Horton,L., Hulme,W., Hiley,I., Johnson,R.,
Jones,C., Karatas,A., Labocque,K., Lamazares,R., Landers,T.,
Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheters,R., Meldrim,J., Menus,L., Mihova,T., Mlenga,Y.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Risse,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schnupack,R., Sedman,S., Severy,P.,
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.

```

```

TITLE
JOURNAL Direct Submission
COMMENT Direct Submission
Submitted (15-DEC-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 23, 2000 this sequence version replaced g1:1181808.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RW/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: MIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1361
Center clone name: 87_N.22

```

```

FEATURES
source 1..193279
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-87N22"
/clone_lib="RP11-87N22 Human Male BAC"
504..561
/rpl_family="(CA)n"
complement(573..882)
/rpl_family="AluY"
988..1015
/rpl_family="(TG)n"
complement(1077..1265)
/rpl_family="MIR"
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region

```

```

repeat_region complement(1319, .1361)
/rpt_family="MIR"
repeat_region complement(1646, .1835)
/rpt_family="MIR"
repeat_region complement(1943, .2320)
/rpt_family="MIR"
repeat_region complement(3451, .3717)
/rpt_family="AluSg"
repeat_region complement(3852, .3981)
/rpt_family="MIR"
repeat_region complement(4678, .4873)
/rpt_family="L1PA3"
repeat_region complement(4907, .4926)
/rpt_family="(A)n"
repeat_region complement(5590, .5758)
/rpt_family="MIR3"
repeat_region complement(7409, .7450)
/rpt_family="T-rich"
repeat_region complement(9037, .9083)
/rpt_family="L2"
repeat_region complement(11416, .11438)
/rpt_family="AT-rich"
repeat_region complement(13115, .13949)
/rpt_family="L2"
repeat_region complement(14968, .15042)
/rpt_family="L2"
repeat_region complement(15058, .15122)
/rpt_family="L2"
repeat_region complement(15282, .15441)
/rpt_family="MIR"
repeat_region complement(15848, .15899)
/rpt_family="(CA)n"
repeat_region complement(17151, .17264)
/rpt_family="MER94"
repeat_region complement(17719, .17739)
/rpt_family="AT-rich"
repeat_region complement(17769, .17839)
/rpt_family="T-rich"
repeat_region complement(18599, .18853)
/rpt_family="L2"
repeat_region complement(21224, .21528)
/rpt_family="AluY"
repeat_region complement(22197, .22307)
/rpt_family="L3"
repeat_region complement(22450, .22875)
/rpt_family="L3"
repeat_region complement(22884, .23008)
/rpt_family="MER20B"
repeat_region complement(23232, .23349)
/rpt_family="L3"
repeat_region complement(25680, .25985)
/rpt_family="AluY"
repeat_region complement(29025, .35170)
/rpt_family="L1PA5"
repeat_region complement(36356, .36375)
/rpt_family="(C)G)n"
repeat_region complement(39370, .39845)
/rpt_family="L1MB3"
repeat_region complement(39876, .40070)
/rpt_family="L1M4"
repeat_region complement(42933, .43060)
/rpt_family="(TA)n"
repeat_region complement(43089, .43144)
/rpt_family="AT-rich"
repeat_region complement(44577, .44696)
/rpt_family="AluY5/8"
repeat_region complement(46470, .46529)
/rpt_family="MIR"
repeat_region complement(47072, .47286)
/rpt_family="MIR"
repeat_region complement(47416, .47698)
/rpt_family="AluSx"
repeat_region complement(49722, .49765)

```

```

repeat_region /rpt_family="(T)G)n"
50251, .50528
/rpt_family="AluYb"
repeat_region complement(55872, .56178)
/rpt_family="AluSx"
repeat_region complement(57210, .57502)
/rpt_family="AluYb"
repeat_region complement(57596, .57678)
/rpt_family="L2"
repeat_region complement(57714, .58028)
/rpt_family="AluSx"
repeat_region complement(58029, .58051)
/rpt_family="AT-rich"
repeat_region complement(58542, .58602)
/rpt_family="MIR3"
repeat_region complement(58932, .59045)
/rpt_family="MER81"
repeat_region complement(59500, .59545)
/rpt_family="MER5A"
repeat_region complement(60676, .60695)
/rpt_family="(A)n"
repeat_region complement(61438, .61631)
/rpt_family="MIR"
repeat_region complement(62350, .62580)
/rpt_family="MIR"
repeat_region complement(62788, .62827)
/rpt_family="AT-rich"
repeat_region complement(63566, .63867)
/rpt_family="L1M4"
repeat_region complement(63934, .64117)
/rpt_family="MIR"
repeat_region complement(64123, .64428)
/rpt_family="AluY5"
repeat_region complement(64779, .65078)
/rpt_family="AluSx"
repeat_region complement(67738, .67813)
/rpt_family="CT-rich"
repeat_region complement(71638, .71784)
/rpt_family="MER5A"
repeat_region complement(71960, .72305)
/rpt_family="L3"

```

```

Query Match 2.08; Score 21; DB 87; Length 193279;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 969 cacccttaacatttgaa 989
Db 9164 CACCTTACATATTGGA 9144
|||||

```

```

RESULT 23
AL451143
LOCUS AL451143 227850 bp DNA HTG 08-APR-2001
DEFINITION Homo sapiens chromosome 6 clone RP11-369A11, *** SEQUENCING IN
PROGRESS ***, 2 unordered pieces.
ACCESSION AL451143
VERSION AL451143.11 GI:13396699
KEYWORDS HTG: HTGS-PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 227850)
AUTHORS Coville,G.
TITLE Direct Submission
JOURNAL Submitted (08-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Mar 20, 2001 this sequence version replaced gi:12963972.
----- Genome Center
Center: Sanger Centre
Center code: SC

```

```

COMMENT

```

Web site: <http://www.sanger.ac.uk>  
 Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
 ----- Project Information  
 Center project name: ba369A11  
 ----- Summary Statistics  
 Assembly program: XGAP4, version 4.5  
 Sequencing vector: M13; M7815; 2% of reads  
 Sequencing vector: plasmid; L08752; 97% of reads  
 Chemistry: Dye-terminator ABI; 0% of reads  
 Chemistry: Dye-terminator Big Dye; 95% of reads  
 Consensus quality: 179880 bases at least Q40  
 Consensus quality: 180269 bases at least Q30  
 Consensus quality: 180455 bases at least Q20  
 Insert size: 227750; sum-of-ctrls  
 Insert size: 170545; 8.7% error; agarose-fp  
 Quality coverage: 12.95x in Q20 bases; sum-of-ctrls quality  
 coverage: 17.29x in Q20 bases; agarose-fp

NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 3437: contig of 3437 bp in length  
 \* 3438 3537: gap of 100 bp  
 \* 3538 227850: contig of 224313 bp in length.

FEATURES  
 source  
 1..227850  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /clone="RP11-369A11"  
 /clone.lib="RPC1-11.2"

misc-feature  
 1..3437  
 /note="assembly\_fragment:00116  
 clone\_end:576  
 vector\_side:left"

misc-feature  
 3538..227850  
 /note="assembly\_fragment:01303  
 clone\_end:77  
 vector\_side:right"

BASE COUNT 69167 a 45830 c 46244 g 66509 t 100 others  
 ORIGIN

Query Match 2.0%; Score 21; DB 81; Length 227850;  
 Best Local Similarity 100.0%; Pred. No. 2.2;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 229 atgtgtatgtgaaggaggtg 249  
 ||||||||||||||||||||  
 Db 183022 ATGTGATGTGAAGGAGTGTG 183042

RESULT 24  
 CEC01G6/c  
 LOCUS CEC01G6 43410 bp DNA INV 23-OCT-2000  
 DEFINITION Caenorhabditis elegans cosmid C01G6, complete sequence.  
 ACCESSION Z3595  
 VERSION Z3595.1 GI:3217055  
 KEYWORDS HTG; 4-commutate-coA ligase; glycerol facilitator protein; Kringle  
 domain; MIP; NADPH-ferrihemoprotein reductase; Protein tyrosine  
 receptor; ror1; Zinc finger (C3HC4 type).  
 SOURCE Caenorhabditis elegans  
 ORGANISM Caenorhabditis elegans  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
 Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.  
 REFERENCE 1 (bases 1 to 43410)  
 AUTHORS none.

TITLE  
 JOURNAL  
 MEDLINE  
 REMARK  
 The C. elegans Sequencing Consortium.  
 Erratum: [[published errata appear in Science 1999 Jan  
 1:283(5398):35 and 1999 Mar 26:283(5410):2103 and 1999 Sep  
 3:285(5433):1493]]  
 2 (bases 1 to 43410)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 Direct Submission  
 Submitted (02-AUG-1994) Nematode Sequencing Project, Sanger Centre,  
 Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,  
 Washington University, St. Louis, MO 63110, USA. E-mail:  
 jes@sanger.ac.uk or rw@nematode.wustl.edu  
 On Jun 13, 1998 this sequence version replaced gi:521002.  
 Coding sequences below are predicted from computer analysis, using  
 predictions from genefinder (P. Green, U. Washington), and other  
 available information.

Current sequence finishing criteria for the C. elegans genome  
 sequencing consortium are that all bases are either sequenced  
 unambiguously on both strands, or on a single strand with both a  
 dye primer and dye terminator reaction, from distinct subclones.  
 Exceptions are indicated by an explicit note.  
 IMPORTANT: This sequence is not the entire insert of clone C01G6.  
 It may be shorter because we only sequence overlapping sections  
 once, or longer because we arrange for a small overlap between  
 neighbouring submissions.

The true left end of clone C01G6 is at 1 in this sequence. The true  
 right end of clone C01G6 is at 42193 in this sequence. The true  
 left end of clone D2013 is at 43257 in this sequence. The start of  
 this sequence (1..104) overlaps with the end of sequence AL021147.  
 The end of this sequence (43257..43410) overlaps with the start of  
 sequence Z47808.  
 For a graphical representation of this sequence and its analysis  
 see: [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=C01G6)  
 name=C01G6.

FEATURES  
 source

1..43410  
 Location/Qualifiers  
 /organism="Caenorhabditis elegans"  
 /db\_xref="taxon:6239"  
 /chromosome="II"  
 /clone="C01G6"  
 complement(join(3007..3031,3318..3476,4137..4605,  
 4867..4993,5528..5620))  
 /gene="C01G6.1a"  
 complement(join(3007..3031,3318..3476,4137..4605,  
 4867..4993,5528..5620))  
 /gene="C01G6.1a"

/note="Similarity with prokaryotic glycerol facilitator  
 proteins belonging to the MIP family, contains similarity  
 to Pfam domain: PF00230 (Major intrinsic protein),  
 Score=151.1, E-value=5.7e-49, N=1  
 CDNA EST EMBL:T01664 comes from this gene; cDNA EST  
 EMBL:T00056 comes from this gene; cDNA EST EMBL:T00055  
 comes from this gene; cDNA EST EMBL:T00161 comes from this  
 gene; cDNA EST EMBL:T00353 comes from this gene  
 CDNA EST EMBL:T00839 comes from this gene; cDNA EST  
 EMBL:T02319 comes from this gene; cDNA EST yk91c9.3 comes from this gene;  
 from this gene; cDNA EST yk91c9.3 comes from this gene;  
 CDNA EST yk96f1.5 comes from this gene; cDNA EST yk61a9.5  
 comes from this gene; cDNA EST yk67a12.5 comes from this gene;  
 comes from this gene; cDNA EST yk80g7.5 comes from this  
 gene; cDNA EST yk80g7.5 comes from this gene; cDNA EST  
 yk85g7.5 comes from this gene  
 CDNA EST yk84a3.5 comes from this gene; cDNA EST yk83b9.5  
 comes from this gene; cDNA EST EMBL:D74401 comes from this  
 gene; cDNA EST yk82h1.5 comes from this gene; cDNA EST  
 yk60c8.5 comes from this gene  
 CDNA EST yk83b9.5 comes from this gene; cDNA EST yk63f1.5  
 comes from this gene; cDNA EST yk80g7.3 comes from this  
 gene; cDNA EST yk85g7.3 comes from this gene; cDNA EST  
 yk96f1.3 comes from this gene

cdna EST EMBL:D71628 comes from this gene; cdna EST  
 yk84a3.3 comes from this gene; cdna EST yk83b9.3 comes  
 from this gene; cdna EST yk60c8.3 comes from this gene;  
 cdna EST yk83b9.3 comes from this gene;  
 cdna EST yk106b4.3 comes from this gene; cdna EST  
 yk67a12.3 comes from this gene; cdna EST  
 yk82a1.3 comes from this gene; cdna EST yk80g1.3 comes  
 from this gene; cdna EST EMBL:T01718 comes from this gene;  
 cdna EST yk63f1.3 comes from this gene; cdna EST yk503q8.3  
 comes from this gene; cdna EST yk505c12.3 comes from this  
 gene; cdna EST yk502f6.3 comes from this gene; cdna EST  
 yk501g10.3 comes from this gene;  
 cdna EST yk493c2.3 comes from this gene; cdna EST  
 yk492h7.3 comes from this gene; cdna EST yk492e.3 comes  
 from this gene; cdna EST yk492e8.5 comes from this gene;  
 cdna EST yk492a1.3 comes from this gene;  
 cdna EST yk492a1.5 comes from this gene; cdna EST  
 yk488f9.3 comes from this gene; cdna EST yk488f9.5 comes  
 from this gene; cdna EST yk488a6.3 comes from this gene;  
 cdna EST yk488a6.5 comes from this gene;  
 cdna EST yk484c1.3 comes from this gene; cdna EST  
 yk484c1.5 comes from this gene; cdna EST yk483e4.3 comes  
 from this gene; cdna EST yk483e4.5 comes from this gene;  
 cdna EST yk479h8.3 comes from this gene; cdna EST  
 yk475d10.3 comes from this gene; cdna EST yk475d10.5 comes  
 from this gene; cdna EST yk464f3.3 comes from this gene;  
 cdna EST yk464f3.5 comes from this gene;  
 cdna EST yk459d12.3 comes from this gene; cdna EST  
 yk459d12.5 comes from this gene; cdna EST yk481d7.3 comes  
 from this gene; cdna EST yk481d7.5 comes from this gene;  
 cdna EST yk470d5.3 comes from this gene; cdna EST  
 yk470a1.3 comes from this gene; cdna EST yk470a1.5 comes  
 from this gene; cdna EST yk466e3.3 comes from this gene;  
 cdna EST yk466e3.5 comes from this gene;  
 yk468e3.3 comes from this gene; cdna EST yk457e11.3 comes  
 from this gene; cdna EST yk457e11.5 comes from this gene;  
 cdna EST yk458g9.3 comes from this gene; cdna EST  
 yk458g9.5 comes from this gene; cdna EST yk456f12.5 comes  
 from this gene; cdna EST yk449b2.3 comes from this gene;  
 cdna EST yk449b2.5 comes from this gene;  
 cdna EST yk435c5.3 comes from this gene; cdna EST  
 yk435c5.5 comes from this gene; cdna EST yk433b8.3 comes  
 from this gene; cdna EST yk433b8.5 comes from this gene;  
 cdna EST yk432g12.3 comes from this gene; cdna EST  
 yk427c11.3 comes from this gene; cdna EST yk427c11.5 comes  
 from this gene; cdna EST yk423g6.3 comes from this gene;  
 cdna EST yk423g6.5 comes from this gene;  
 cdna EST yk423c10.3 comes from this gene; cdna EST  
 yk423c10.5 comes from this gene; cdna EST yk425g5.3 comes  
 from this gene; cdna EST yk425g5.5 comes from this gene;  
 cdna EST yk425d2.3 comes from this gene;  
 cdna EST yk425d2.5 comes from this gene; cdna EST  
 yk441f6.3 comes from this gene; cdna EST yk414f8.5 comes  
 from this gene; cdna EST yk416c12.3 comes from this gene;  
 cdna EST yk416c12.5 comes from this gene;  
 cdna EST yk416b1.3 comes from this gene; cdna EST  
 yk416b1.5 comes from this gene; cdna EST yk400h1.3 comes  
 from this gene; cdna EST yk400h1.5 comes from this gene;  
 cdna EST yk400c10.3 comes from this gene;  
 cdna EST yk400c10.5 comes from this gene; cdna EST  
 yk39g05.3 comes from this gene; cdna EST yk39g45.5 comes  
 from this gene; cdna EST yk393a10.3 comes from this gene;  
 cdna EST yk393a10.5 comes from this gene;  
 cdna EST yk376e2.3 comes from this gene; cdna EST  
 yk376e2.5 comes from this gene; cdna EST yk373e9.3 comes  
 from this gene; cdna EST yk373e9.5 comes from this gene;  
 cdna EST yk373c11.3 comes from this gene; cdna EST  
 yk373c11.5 comes from this gene; cdna EST

yk367b6.3 comes from this gene; cdna EST yk367b6.5 comes  
 from this gene; cdna EST yk465e6.3 comes from this gene;  
 cdna EST yk465e6.5 comes from this gene;  
 cdna EST yk343e3.3 comes from this gene; cdna EST  
 yk343e3.5 comes from this gene; cdna EST yk340d8.3 comes  
 from this gene; cdna EST yk340d8.5 comes from this gene;  
 cdna EST yk328g4.3 comes from this gene;  
 cdna EST yk328g4.5 comes from this gene; cdna EST  
 yk333d2.3 comes from this gene; cdna EST yk333d2.5 comes  
 from this gene; cdna EST yk334h10.3 comes from this gene;  
 cdna EST yk334h10.5 comes from this gene;  
 cdna EST yk332e3.3 comes from this gene; cdna EST  
 yk332e3.5 comes from this gene; cdna EST yk329a10.3 comes  
 from this gene; cdna EST yk329a10.5 comes from this gene;  
 cdna EST yk319e8.3 comes from this gene;  
 cdna EST yk319e8.5 comes from this gene; cdna EST  
 yk314h11.3 comes from this gene; cdna EST yk314h11.5 comes  
 from this gene; cdna EST yk312c9.3 comes from this gene;  
 cdna EST yk312c9.5 comes from this gene;  
 cdna EST yk289e2.3 comes from this gene; cdna EST  
 yk289e2.5 comes from this gene; cdna EST yk281d3.3 comes  
 from this gene; cdna EST yk281d3.5 comes from this gene;  
 cdna EST yk245g6.3 comes from this gene;  
 cdna EST yk245g6.5 comes from this gene; cdna EST  
 yk257b5.3 comes from this gene; cdna EST yk257b5.5 comes  
 from this gene; cdna EST yk234e3.3 comes from this gene;  
 cdna EST yk234e3.5 comes from this gene;  
 cdna EST yk246d1.3 comes from this gene; cdna EST  
 yk229g12.3 comes from this gene; cdna EST yk229g12.5 comes  
 from this gene; cdna EST yk213e7.3 comes from this gene;  
 cdna EST yk213e7.5 comes from this gene;  
 cdna EST yk198g10.3 comes from this gene; cdna EST  
 yk198g10.5 comes from this gene; cdna EST yk313h9.5 comes  
 from this gene; cdna EST yk313h9.3 comes from this gene;  
 cdna EST yk333e2.5 comes from this gene;  
 cdna EST yk333e2.3 comes from this gene; cdna EST  
 yk459g1.5 comes from this gene; cdna EST yk477e6.5 comes  
 from this gene; cdna EST yk463d2.3 comes from this gene;  
 cdna EST yk463g2.3 comes from this gene;  
 cdna EST yk510g4.3 comes from this gene; cdna EST  
 yk511h8.3 comes from this gene; cdna EST yk512c3.3 comes

Query Match 1.9%; Score 20; DB 5; Length 43410;  
 Best Local Similarity 100.0%; Pred. No. 8.4;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1027 acaaacgaagattcgat 1046  
 ||||||||||||||||  
 Db 38560 ACAACCGAGAGATTGCAAT 38541

RESUT 25  
 AC005142/c DNA 17-FEB-1999  
 LOCUS Arabidopsis thaliana BAC T5L23 from chromosome IV, near 19 cM,  
 DEFINITION complete sequence.

ACCESSION AC005142  
 VERSION AC005142.2 GI:4263038  
 KEYWORDS HTG.  
 SOURCE  
 ORGANISM

thale cress.  
 Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliopsida; eudicotyledons; core eudicots; Rosidae; eurosids II;

Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 116448)  
 AUTHORS Zhong,J., Ma,P., Parnell,L.D., Chen,C.-N. and Chen,E.Y.

TITLE BAC T5L23 from chromosome IV, position 19 cM  
 JOURNAL Unpublished

REFERENCE 2 (bases 1 to 116448)  
 Zhong,J., Ma,P., Parnell,L.D., Chen,C.-N. and Chen,E.Y.

AUTHORS Direct Submission  
 TITLE Submitted (19-JUN-1998) Applied Biosystems Division of  
 JOURNAL Perkin-Elmer, 850 Lincoln Centre Dr., Foster City, CA 94404, USA

REFERENCE 3 (bases 1 to 116448)  
 AUTHORS Parnell,L.D. and Chen,E.Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-FEB-1999) Applied Biosystems Division, Perkin Elmer Corporation, 850 Lincoln Centre Dr., Foster City, CA 94404  
 REMARK Arabidopsis thaliana BAC T5L23 from chromosome IV near 19 cM  
 On Feb 22, 1999 this sequence version replaced g1:3241935.  
 BAC T5L23 maps to near 19 cM on the Lister & Dean RI map and is assigned to YAC C1C8B1. Position 1 of T5L23 is oriented toward the telomere and position 116473 is oriented toward the centromere. For more information on the mapping, sequencing and annotation of T5L23, please see <http://www.cshl.org/arabweb/T5L23-titlepage.html>.  
 A graphic view of our annotation is also available at this url.  
 Gene models are built with exons predicted by GenScan  
 (<http://CCR-081.mit.edu/GENSCAN.html>), MZEF  
 (<http://www.cshl.org/genefinder>) and GRAIL  
 (<http://compbio.ornl.gov/tools/index.shtml>) and with splice sites predicted by NetPlantGene  
 (<http://www.cbs.dtu.dk/netgene/cbsnetgene.html>). Genes are numbered according to the scheme BAC.gene.number. Typically, these numbers progress from 1 upwards as one moves from position 1 of the BAC. Protein sequences encoded by the genes are assigned to a functional category with the aid of similarity searches and comparison to the Prosite  
 (<http://expasy.hugue.ch/sprot/prosite.html>) and Pfam  
 (<http://pfam.wustl.edu/>) libraries. A description of these categories can be found at  
<http://muntjac.mips.biochem.mpg.de/Arabidopsis/>. Genomic repeats are typically located by TBLASTX analysis and an attempt is made to classify the function of each repeat as either transposon, putative microsatellite, LINE, direct repeat, centromeric repeat, etc.

If you have any questions or confirmatory or contradictory evidence concerning the annotation of T5L23, please direct email to Larry Parnell at [parnell@cschl.org](mailto:parnell@cschl.org).

#### FEATURES

##### SOURCE

1. 116448  
 /organism="Arabidopsis thaliana"  
 /cultivar="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="IV"  
 /map="19 cM"  
 /note="hybridizing YAC is C1C8B1"

##### misc\_feature

1. 31783  
 /note="overlap with F9H3, GenBank accession number AF071527, from position 1 to 31783"  
 /note="Region: overlap with F9H3"  
 complement(2025..3495)  
 /gene="T5L23.1"

##### gene

/note="encodes putative M-type thiorodoxin; identical to F9H3.15, GenBank accession number AF071527; genomic copy of EST AA042639; genomic copy of EST AA650755; gene model last edited on 5 Jan 99"  
 /evidence-not\_experimental  
 complement(join(2025..2341,3252..3495))  
 /gene="T5L23.1"

##### CDS

/note="similar to A. thaliana protein F21B7.7, GenBank accession number AC002560; M-type thiorodoxin is also present in cyanobacteria and chloroplasts; functional catalog ID=02"  
 /codon\_start=1  
 /evidence-not\_experimental  
 /product="putative M-type thiorodoxin"  
 /protein\_id="AADI5308.1"  
 /db\_xref="GI:4263039"  
 /translation="MAAFTCTSRPPIISLSETRIVSSPSSASSLRMPAVLPSSG  
 LRRLSPASLTIHOPRVSRLRAVVEAQTETDIOVNDSTWDSLVKATGPV  
 VDMAPGCPKMDLPVNDLAQHYGKIKFKINTDESNTPGOYGVRSIPITMIV  
 GGRKDTIIGAVPKTLLISDKELP"  
 complement(4139..4163)  
 /note="function=putative\_microsatellite"  
 /rpt\_type=tandem

##### repeat\_region

/rpt\_unit=A  
 5120..5124  
 /note="function=terminal\_flanking\_direct\_repeat; cognate A of flanking direct repeat of LTR transposon from position 5145 to 6873; cognate B resides from position 6869 to 6873"  
 /rpt\_type=flanking

##### repeat\_region

5120..6848  
 /note="function=LTR\_transposon"  
 /rpt\_type=dispersed

##### LTR

5125..5328  
 /note="cognate A of LTR flanking transposon from position 5145 to 6873; cognate B resides from 6665 to 6868"  
 5508..6632  
 /gene="T5L23.2"

##### gene

/note="encodes putative reverse transcriptase; likely functions as an element of LTR transposon from position 5145 to 6873; gene model last edited on 5 Jan 99"  
 /evidence-not\_experimental  
 5508..6632  
 /gene="T5L23.2"

##### CDS

/note="likely element of LTR transposon from position 5145 to 6873"  
 /codon\_start=1  
 /evidence-not\_experimental  
 /product="putative reverse transcriptase"  
 /protein\_id="VAD15309.1"

/db\_xref="GI:4263040"  
 /translation="MGRKMGKSVRRSSPDLLVIAASKIHBOSSGSSARAPSSIIHA  
 IHOPAYSSRHEVEDSPDSSHPYHLVSSDHPIGLVLAPELDGNSYGTWIIAMTS  
 IEAKKLGPEFGSIPKDDDDPYCKIWRRCNSMVKSWILNSVSKETIYSLFPTAAA  
 IMLKTLRPFHKSLLPRLYLRQIHLRQGNLDLSYHRRQOTLMEETLSAIPRTV  
 EBLLETRTNRYTDLMLGNCYDVAVRSOILMKTLPLSLFPMIDODEIORSARIS  
 TRPGWTSVFRVNSQSSQSVLNGDTYOKRKRPYCYGSPRHVEDTQYKKHGGIPSPFK  
 SKQINMELQIVDFETKLIFSHQRLSKMGLNITVPS"

##### LTR

6640..6843  
 /note="cognate B of LTR flanking transposon from position 5145 to 6873; cognate A resides from 5150 to 5353"

##### repeat\_region

6844..6848  
 /note="function=terminal\_flanking\_direct\_repeat; cognate B of flanking direct repeat of LTR transposon from position 5145 to 6873; cognate A resides from position 5145 to 5149"

##### repeat\_region

/rpt\_type=flanking  
 complement(7885..7904)  
 /note="function=putative\_microsatellite"

##### repeat\_region

/rpt\_type=tandem  
 complement(7905..8818)  
 /note="function=unclassified; similar to MX110, GenBank accession number AB005248"  
 /rpt\_type=dispersed  
 9482..10923  
 /gene="T5L23.3"

##### gene

/note="encodes hypothetical protein; identical to F9H3.17, GenBank accession number AF071527; gene model last edited on 5 Jan 99"  
 /evidence-not\_experimental  
 join(9482..9639,9745..9847,10690..10923)  
 /gene="T5L23.3"

##### CDS

/note="identical to F9H3.17, GenBank accession number AF071527; similar to F21B7.5, GenBank accession number AC002560; functional catalog ID=99"  
 /codon\_start=1  
 /evidence-not\_experimental  
 /product="hypothetical protein"

/protein\_id="AADI5310.1"  
 /db\_xref="GI:4263041"

/translation="WVKLRIGGLVLRARGAALAIWTTSREPASLATSLEA  
 KYTDMARFETVIANAVSVSLVLPKRSILMKFVYVLDVLTWMLTSSSAAIA  
 VAOVGKRGKNNANWMLPTGQVPRFCDITGALVIALVLTLLTSLRAVDPPL  
 LQKS"  
 12953..18635

##### gene

CDS

```

/gene="75L23.4"
/note="encodes putative glucan synthase component;
identical to F9H3.18, GenBank accession number AF071527;
gene model last edited on 5 Jan 99"
/evidence=not-experimental
join(12953..14823,14984..17006,17187..18635)
/gene="75L23.4"
/note="similar to 1,3-beta glucan synthase; identical to
F9H3.18, GenBank accession number AF071527; functional
catalog ID=01.05.99; functional catalog ID=09.01"
/codon_start=1
/evidence=not-experimental
/product="putative glucan synthase component"
/protein_id="AA015311.1"
/db_xref="GI:4263042"
/translation="MSLRHTVPQTRPLAEAVGIEEPYNIIPVNNLADPSLR
PEVRAAALATVGDLRPPYQWRSHYDLMLALFPGKQDVNRQREHWHLIA
NMOMRISPPNDIDSLDSAVRFRKRLANYSWCSYGGKSNWISDRNPSRREL
LVYGLYLIMGEANLRFMEPCICYIFNNASSELNKLDECDENTGOPYLPSLGEN
AEUTGVKPIYDTIOAEIDSKNGTVAHCKWRNTDINPEPTDCSEKLKWPDLGS
NFKSRGKSVGKGTGFEVERFPFYLYRSPDRPLWVLAFLQAAIYAMEEKPTSSVTR
OLMNAIKAROVRLIPLVLTWSGRRLQANTLHAQSTPLVSRKTRFPMKLVIA
AAVWIVATVYLYTNIKOKRODRONSNAITKTYPLVAVGAPLPELLALFIIPW
MNPRLLEPNKTIFFALTMFQGRSVGRLREGLVNIKYSTFWIVLATKFESEYL
YKAPMKPSKLLMNLADVDYEWHPQSDNSFVALMLVLLVLYLMDIOWAITS
IVGAVVGLDHLGEIRDMQGLRFRQFASIOFNLMPEQLNARGKPKFGDTHR
LKLRFGRPLCNEELLASQARELIDADPKMLHCKREYRCAVEADSTKHLIL
SLIKVDTEHSIIIVFEOIINOSIOSEQPTFRVLLPKIYETLOKVCYLVDEEDT
SGRVNVLQSLIYELATROPEIKTTRTNEGLTPRDPASKLFPQNTLPSPASMD
FYROYRRLHTLITRDSMSHSPVNLKARRIATFSSLPNMPHAPVEMKAFSVLT
PYSEEVYKREOIRNETEDGISTYLYLQTTIADENKNEKMHREGIKTDSLEWTK
LIDLILNASTRGOTLARTYRGMMTYRRLMFLQDSASEMDIREGQELGYSVRLNIG
ELGGSODEFVSENDSSLSRASSVSITLYKGHEVGTALMKFVYVACQIYGSOKAKE

```

Query Match 1.9%; Score 20; DB 12; Length 116448;  
 Best Local Similarity 100.0%; Pred. No. 8.6;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1031 accgaagatcgaatgcatg 1050  
 |||||||||||||||||||

Db 82946 ACCGAGAGATTGCAATGAT 82927

RESULT 26  
 AC011195 LOCUS  
 DEFINITION  
 AC011195.9 GI:13606020  
 VERSION  
 AC011195.9  
 KEYWORDS  
 HMG; HMGs\_PHASEL.  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 147620)  
 Birten, B., Linton, L., Nusbaum, C. and Lander, E.  
 Homo sapiens chromosome 17, clone CTD-2200P10  
 Unpublished  
 2 (bases 1 to 147620)  
 Birten, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,  
 Balwin, J., Barna, N., Becker, R., Boguslavsky, L., Boukhalter, B.,  
 Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,  
 Cooke, P., DeRellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,  
 Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,  
 Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
 Howland, J., Johnson, R., Jones, C., Kann, B., Karas, A., Klein, J.,  
 Leinoczy, J., Lieu, C., Locke, K., MacDonald, P., Marquis, N.,  
 McEwan, P., McGurk, A., McKernan, K., McDonald, P., Meldrum, J.,  
 Morrow, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
 Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,

TITLE  
 JOURNAL  
 COMMENT

Testafay, S., Tirtrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,  
 Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (01-OCT-1999) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr. 12, 2001 this sequence version replaced gi:3489206.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

Project Information

Center project name: 1865

Center clone name: 2200\_P\_10

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 13 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 9600: contig of 9600 bp in length  
 9601 9700: gap of 100 bp  
 9701 70619: contig of 60919 bp in length  
 70620 70719: gap of 100 bp  
 70720 73046: contig of 2327 bp in length  
 73047 73146: gap of 100 bp  
 73147 77498: contig of 4352 bp in length  
 77499 77598: gap of 100 bp  
 77599 82519: contig of 4821 bp in length  
 82520 82619: gap of 100 bp  
 82620 95179: contig of 12560 bp in length  
 95180 95279: gap of 100 bp  
 95280 104686: contig of 9407 bp in length  
 104687 104786: gap of 100 bp  
 104787 129213: contig of 24427 bp in length  
 129214 129313: gap of 100 bp  
 129314 131809: contig of 2596 bp in length  
 131910 132009: gap of 100 bp  
 132010 136278: contig of 4269 bp in length  
 136279 136378: gap of 100 bp  
 136379 139944: contig of 3566 bp in length  
 139945 140044: gap of 100 bp  
 140045 144250: contig of 4206 bp in length  
 144251 144350: gap of 100 bp  
 144351 147620: contig of 3270 bp in length.

FEATURES  
 SOURCE

1. 147620  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="17"  
 /map="17"  
 /clone="CTD-2200P10"  
 /clone\_11b="CTD Human BAC"  
 BASE COUNT 41373 a 33542 c 32455 g 38953 t 1297 others  
 ORIGIN

Query Match 1.9%; Score 20; DB 62; Length 147620;  
 Best Local Similarity 100.0%; Pred. No. 8.6;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 413 ggaagcaactcgtgttt 432  
 |||||||||||||||||||

Db 65825 GGAAGCAACTCAGTGT 65844

RESULT 27



AC009445/c  
 LOCUS AC009445 149064 bp DNA HTG 10-SEP-2000  
 DEFINITION Homo sapiens clone RP11-1P8, WORKING DRAFT SEQUENCE, 15 unordered  
 pieces.  
 AC009445  
 AC009445.3 GI:10047682  
 VERSION AC009445.3  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 149064)  
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
 Homo sapiens, clone RP11-1P8  
 Unpublished  
 2 (bases 1 to 149064)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,  
 Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A.,  
 Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A.,  
 Cooke, P., DeRianno, K., Depayre, E., Devon, K., Dewar, K.,  
 Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C.,  
 Funke, R., Gage, D., Galagan, J., Gardy, S., Gilbert, D., Grant, G.,  
 Hagos, B., Heath, A., Horton, L., Howland, J., Jones, C., Kam, L.,  
 Karas, A., Lehoczy, J., Lien, C., Locke, K., Macdonald, P.,  
 Margis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J.,  
 Melidrim, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,  
 Naylor, J., Nijhoff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,  
 Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,  
 Strange-Thomann, N., Stojanovic, N., Stone, C., Subramaniam, A.,  
 Testa, S., Toriella-Waller, I., Vassiliou, H., Vo, A., Wagner, A.,  
 Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.  
 Direct Submission  
 Submitted (22-AUG-1999) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Sep 10, 2000 this sequence version replaced gi:7622308.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 -----  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WITB  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 -----  
 Project Information  
 Center project name: 1.P.8  
 Center clone name: 1.P.8  
 -----  
 Summary Statistics  
 Sequencing vector: M13; M77815; 100% of reads  
 Chemistry: Dye-primer-amersham; 6% of reads  
 Chemistry: Dye-terminator Big Dye; 94% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 136415 bases at least Q40  
 Consensus quality: 141840 bases at least Q30  
 Consensus quality: 14446 bases at least Q20  
 Insert size: 154000; agarose-mp  
 Insert size: 147664; sum-of-contents  
 Quality coverage: 3.9 in Q20 bases; agarose-mp  
 Quality coverage: 4.1 in Q20 bases;  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 15 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1  
 \* 4988: contig of 4988 bp in length  
 \* 4989 5088: gap of 100 bp  
 \* 5089 6190: contig of 1102 bp in length  
 \* 6191 6290: gap of 100 bp  
 \* 6291 9698: contig of 3408 bp in length  
 \* 9699 9798: gap of 100 bp  
 \* 9799 14363: contig of 4565 bp in length

FEATURES  
 Source  
 \* 14364 14463: gap of 100 bp  
 \* 14464 43843: contig of 29380 bp in length  
 \* 43844 43943: gap of 100 bp  
 \* 43944 50139: contig of 6196 bp in length  
 \* 50140 50239: gap of 100 bp  
 \* 50240 58176: contig of 7937 bp in length  
 \* 58177 58276: gap of 100 bp  
 \* 58277 66999: contig of 8723 bp in length  
 \* 67000 67099: gap of 100 bp  
 \* 67100 76323: contig of 9224 bp in length  
 \* 76324 76423: gap of 100 bp  
 \* 76424 83260: contig of 6837 bp in length  
 \* 83261 83360: gap of 100 bp  
 \* 83361 93552: contig of 10192 bp in length  
 \* 93553 93652: gap of 100 bp  
 \* 93653 103776: contig of 10124 bp in length  
 \* 103777 103876: gap of 100 bp  
 \* 103877 118277: contig of 14401 bp in length  
 \* 118278 118377: gap of 100 bp  
 \* 118378 144219: contig of 25842 bp in length  
 \* 144220 144319: gap of 100 bp  
 \* 144320 149064: contig of 4745 bp in length.  
 Location/Qualifiers  
 1. 149064  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="RP11-1P8"  
 /clone\_1fb="RPC1-11 Human Male BAC"  
 1. 4988  
 /note="assembly-fragment  
 clone\_end:sp6  
 vector\_side:left"  
 misc\_feature  
 5089..6190  
 /note="assembly-fragment"  
 misc\_feature  
 6291..9698  
 /note="assembly-fragment"  
 misc\_feature  
 9799..14363  
 /note="assembly-fragment"  
 misc\_feature  
 14464..43843  
 /note="assembly-fragment"  
 misc\_feature  
 43944..50139  
 /note="assembly-fragment"  
 misc\_feature  
 50240..58176  
 /note="assembly-fragment"  
 misc\_feature  
 58277..66999  
 /note="assembly-fragment"  
 misc\_feature  
 67100..76323  
 /note="assembly-fragment"  
 misc\_feature  
 76424..83260  
 /note="assembly-fragment"  
 misc\_feature  
 83361..93552  
 /note="assembly-fragment"  
 misc\_feature  
 93653..103776  
 /note="assembly-fragment"  
 misc\_feature  
 103877..118277  
 /note="assembly-fragment"  
 misc\_feature  
 118378..144219  
 /note="assembly-fragment"  
 misc\_feature  
 144320..149064  
 /note="assembly-fragment  
 clone\_end:77  
 vector\_side:right"  
 BASE COUNT 47071 a 28109 c 27442 g 45042 t 1400 others  
 ORIGIN  
 Query Match 1.9% Score 20; DB 61; Length 149064;  
 Best Local Similarity 100.0%; Pred. No. 8.6;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 734 aaaaatttcagatgcacatc 753  
 |||||  
 Db 134850 AAAAGTTTCAGATGCACATG 134831

RESULT 28  
AL512606  
LOCUS Homo sapiens chromosome 9 clone RP11-1P8, \*\*\* SEQUENCING IN  
DEFINITION  
ACCESSION AL512606.2 GI:13561312  
VERSION AL512606.2 GI:13561312  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 153097)  
AUTHORS Plumb, B.  
JOURNAL Direct Submission  
Submitted (06-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
On Apr 8, 2001 this sequence version replaced gi:12193194.  
Draft Sequence produced by Whitehead Institute/MIT Center for  
Genome Research 320 Charles Street,  
Cambridge, MA 02141, USA  
http://www.seq.wi.mit.edu  
----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: DALP8  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: M13; M77815; 59% of reads  
Chemistry: Dye-terminator Big Dye; 96% of reads  
Chemistry: Dye-terminator Big Dye; 96% of reads  
Consensus quality: 150523 bases at least Q40  
Consensus quality: 151584 bases at least Q30  
Insert size: 152697; sum-of-contigs  
Insert size: 149419; 7.4% error; agarose-fp  
Quality coverage: 6.19x in Q20 bases; sum-of-contigs quality  
coverage: 6.33x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 5 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1 15753: contig of 15753 bp in length  
\* 15754 15853: gap of 100 bp  
\* 15854 50145: contig of 34292 bp in length  
\* 50146 50245: gap of 100 bp  
\* 50246 110207: contig of 59962 bp in length  
\* 110208 110307: gap of 100 bp  
\* 110308 130603: contig of 20296 bp in length  
\* 130604 130703: gap of 100 bp  
\* 130704 153097: contig of 22394 bp in length.  
Location/Qualifiers  
1. 153097  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="9"  
/clone="RP11-1P8"  
/clone\_lib="RPC1-11.1"  
1. 15753  
/note="assembly\_fragment:01279  
clone\_end:SP6

misc\_feature  
15854..50145  
/note="assembly\_fragment:01208  
fragment\_chain:1"  
50246..110207  
/note="assembly\_fragment:00944  
fragment\_chain:1"  
110308..130603  
/note="assembly\_fragment:01663  
fragment\_chain:2"  
130704..153097  
/note="assembly\_fragment:00796  
fragment\_chain:2  
clone\_end:T7  
vector\_side:right"  
BASE COUNT 47667 a 28985 c 28429 g 47613 t 403 others  
ORIGIN

Query Match 1.9%; Score 20; DB 81; Length 153097;  
Best Local Similarity 100.0%; Pred. No. 8.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 734 aaagttgcagatgacatg 753  
|||||  
DB 53019 AAAAGTTGCAGATGACATG 53038

RESULT 29  
ATCHRIV9/c  
LOCUS Arabidopsis thaliana DNA chromosome 4, contig fragment No. 9.  
DEFINITION  
ACCESSION AL161497  
VERSION AL161497.2 GI:7270670  
KEYWORDS  
SOURCE  
ORGANISM  
thale cress.  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 21279 to 137726)  
REFERENCE  
AUTHORS Zhong, J., Ma, P., Parnell, L. D., Chen, C. N., Chen, E. Y., Mewes, H. W.,  
Lemcke, K., and Mayer, K. F. X.  
JOURNAL Unpublished  
REFERENCE  
AUTHORS 2 (bases 137729 to 137730)  
Lammer, B., Stoneking, T., Stumpf, J., Mewes, H. W., Lemcke, K., and  
Mayer, K. F. X.  
JOURNAL Unpublished  
REFERENCE  
AUTHORS 3 (bases 1 to 53060)  
Spiegel, L. A., Huang, E. N., Nascimento, L. U., de la Bastide, M.,  
Vil, D. M., Preston, R. R., Matero, A., Shah, R., O'Shaughnessy, A.,  
Rodriguez, M., Shekher, M., Schutz, K., See, L. H., Sady, I.,  
Habermann, K., Dedhia, N. N., Mewes, H. W., Lemcke, K., and Mayer, K. F. X.  
JOURNAL Unpublished  
REFERENCE  
AUTHORS 4 (bases 1 to 159629)  
EU Arabidopsis sequencing project.  
JOURNAL Direct Submission  
Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer  
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG. E-mail:  
lemcke@mps.biochem.mpg.de, mayer@mps.biochem.mpg.de Project  
Coordinator: Mike Beyer, Molecular Genetics Department, Cambridge  
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,  
E-mail: michael.bevan@bsrc.ac.uk  
Information on performance of analysis and a more detailed  
annotation of this entry and other sequences of chromosomes 3, 4  
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/  
this fragment has an overlap with ATCHRIV8 at the 5' end and an  
overlap with ATCHRIV10 at the 3' end.  
Location/Qualifiers  
1. 159629  
/organism="Arabidopsis thaliana"  
/variety="Columbia"  
/db\_xref="taxon:3702"

## FEATURES

source

misc\_feature

1. 15753  
/note="assembly\_fragment:01279  
clone\_end:SP6

gene /chromosome="4"  
complement(join(6741..7064,7161..7708,7783..8101,  
8409..8866,9283..9343,9445..9588))  
/gene="AT4g03480"  
exon complement(6741..7064)  
/gene="AT4g03480"  
number=1  
741..9588  
/gene="AT4g03480"  
complement(join(6741..7064,7161..7708,7783..8101,  
8409..8866,9283..9343,9445..9588))  
/gene="AT4g03480"  
/note="contains similarity to mammalian ankyrins  
similarity to similar to A. thaliana hypothetical protein  
F2187.8, GenBank accession number AC002560  
contains ATP/GTP-binding site motif A (P-loop) AA297-304  
contains EST gb:H37081, T42092"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAB7783.1"  
/db\_xref="GI:7270671"  
translation="MWRSHLRATPEASPIKPASSPNCRLSLSPVELRLYPIRDDF  
KYKIDIKTSYONSLFELPCDSEKERTLSRSDLPNPGEPYPMPELFSAMRGNI  
ELKLEKSTETPMACKSGDSVYLHAASGHELVKNITTEPCILLPNSKQIIP  
LHAARAGSAVVKALVASLYFSPRVEEDRLNITVYLDIDGPIPLAAKLDHE  
KAERIRKLSLSHLIMHMRSRCSFSDASTROMETACLVNADQHASFTLAKDSTP  
LYLAEGANVSLVRAMLNPGNKIQGKSTLSOLEGRKSLHAALAKAKTDLNVL  
NDPSLVNERDEGRCLSVGASMGYKKGICLLRSTSVYECDDGSGPFIHMAVEK  
GHLKVAKEILKRCPSKELVKNOCOMLHAAKSAKVSFLLGYIRLDELNHLLEEG  
DVGDNAPIELATINMRCRTRVDKLAAPFSTETKILNOKMDGLRDLDAELNLOPYVL  
REELTILWILCYAPKSYGALPTSGMTLRSRSPDAAKKYKDHINALLIATVATVI  
FPAGFTTIPGCFNSSAPNMKATLADDDSTLEFVLDTLAMOSSTIVALI"1"  
complement(7065..7160)  
number=1  
complement(7161..7708)  
/gene="AT4g03480"  
number=2  
complement(7709..7782)  
/number=2  
complement(7783..8101)  
/gene="AT4g03480"  
number=3  
complement(8102..8408)  
/number=3  
complement(8409..8866)  
/gene="AT4g03480"  
number=4  
complement(8867..9282)  
/number=4  
complement(9283..9343)  
/gene="AT4g03480"  
number=5  
complement(9344..9444)  
/number=5  
complement(9445..9588)  
/gene="AT4g03480"  
number=6  
complement(10335..10525)  
/gene="AT4g03490"  
number=1  
complement(join(10335..10525,10983..11078,11590..12133,  
12345..12500,12627..12772,12871..13501))  
/gene="AT4g03490"  
10335..13501  
/gene="AT4g03490"  
complement(join(10335..10525,10983..11078,11590..12133,  
12345..12500,12627..12772,12871..13501))  
/gene="AT4g03490"  
12345..13501  
/gene="AT4g03490"  
complement(join(10335..10525,10983..11078,11590..12133,  
12345..12500,12627..12772,12871..13501))  
/gene="AT4g03490"  
/note="contains similarity to mammalian ankyrins  
similarity to similar to A. thaliana hypothetical protein  
F2187.8, GenBank accession number AC002560"  
/codon\_start=1  
/product="hypothetical protein"

/protein\_id="CAB7783.1"  
/db\_xref="GI:7270672"  
translation="MGKVOEFKEKMEENEIPVLQDVFQGNITLILAIYGHDLVR  
ILAEELNILLNMKRLGNLNCNFPSPFSHYOTLVARRYKODLHVAAAAGHKLIYGLLI  
DLAELOLFDITWVIGSEQMVIGNIFRYSNNQDNTALHLSLKGNHVSLSQDLVREDRST  
CEFLDKEDVSPLYMAAEAGVYSILVEHMLRGLDASVCKSVCAAYSOQLDLTAAYE  
SDSDLVESNDEDEGRTPLATNASIGYDIGOYHMLTFEASSTOGONVILVNAKSGNARAY  
GYILRKSDYKRLINEODTEGRTPLHASSNRPKYLIMALVAAGTTPAPRYLREID  
IPGLTTDEDILIKIRKDRNTLLVATLVATMAFAAGLSPLPGYNSTEEKSVYKSYE  
ESATFAVJCNLSIAVYTAIVSTVALIGTQDLADKCOMLTERETFPILGFSITANSLAE  
VAGLYLVIGHHWLAIFVLASGGFTLMALLLIIPYASPTFTLSLSLVSQNNSE  
DVDSVNOLVAPTEELALIKAIQVNRIOGNIGVTSLSYLLKRVAPINVLLHLKLN  
HSLVFSGHLYDGPKKYTSOL"  
complement(10526..10982)  
number=1  
complement(10983..11078)  
/gene="AT4g03490"  
number=2  
complement(11079..11589)  
/number=2  
complement(11590..12133)  
/gene="AT4g03490"  
number=3  
complement(12134..12344)  
/number=3  
complement(12345..12500)  
/gene="AT4g03490"  
number=4  
complement(12501..12626)  
/number=4  
complement(12627..12772)  
/gene="AT4g03490"  
number=5  
complement(12773..12870)  
/number=5  
complement(12871..13501)  
/gene="AT4g03490"  
number=6  
13514..15185  
/note="marker m1233"  
14170..17288  
/gene="AT4g03500"  
join(14170..14204,15103..15848,15939..16477,16650..17288)  
/gene="AT4g03500"  
/note="hypothetical protein with some similarity to  
ankyrin  
similarity to"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAB7783.1"  
/db\_xref="GI:7270673"  
translation="MKNESRTSEELSLGIFSGVGSMDRPVHLDRRAENHQAIPIR  
NNVRSSTSLDSTLPDFTSETKPMDPKPTMAAVRAGENVLSNNSYISVAPTLVNDRG  
NTILHAASGHVSLVRYITCKPGLLKNMGEVALHLAAEGHLDVYWNLIIDFPI  
DISCTNLPVAKRIYFAKKNODTALHVALKGHEVAVLSYSAKSLSPFANNDLSEF  
LYLAEGHRTSLVITMCHGTNETSSKYGSRIVYHAALANKRKDILLDLAKSALINL  
RDEGRTISLFGASIGYQGFSTLPDKNRDQVYVSDDCLPFHMAAYGQVOTLEETL  
KHCPEALBLDRQNTLHLAKTGKLVKIFTLSCCKKKKLLINBQVNGNTPILH  
LATINMHPKXVSMFTWDRHVDLKRNTIGTDLVAEENIDISTIVHORLMMALINA  
GAPKSTPTENLRSFKKPDGSKYKDRVNTLMLVATLVATMTPTAGTTLPGGYNDSFP  
HIGMAVLAKTAFQVAVLVCDTLAWSSITITVALTMAQIGDSTILKAFNIALPELGL  
ALTSMSIAFMAGTYVAVSHPLPGYFVLGIGITFILLVLLLVVYSPVMAAPLRLH  
IFYYPFLKLAAGDKNKIVDYTTASDE"  
14170..14204  
/gene="AT4g03500"  
number=1  
14205..15102  
/gene="AT4g03500"  
number=1  
15103..15848  
/gene="AT4g03500"  
number=2  
15849..15938

intron

exon /gene="AT4g03500"  
/number=2  
15939..16477  
/gene="AT4g03500"  
/number=3  
16478..16649  
/gene="AT4g03500"  
/number=3  
16650..17288  
/gene="AT4g03500"  
/number=4

Query Match 1.9% Score 20; DB 13; Length 159629;  
Best Local Similarity 100.0%; Pred. No. 8.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1031 accgaagattcgaatgat 1050  
|||||  
Db 104224 ACCAGAGATTTCGATGAT 104205

RESULT 30  
AL390882/c 162064 bp DNA PRI 28-FEB-2001  
LOCUS Human DNA sequence from clone RP11-113D19 on chromosome  
DEFINITION 9p21.2-22.3, complete sequence.  
ACCESSION AL390882  
VERSION AL390882.12 GI:12657180  
KEYWORDS HMG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 162064)  
Donnelly,S.  
Direct Submission  
Submitted (28-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
Requests: clonerequest@sanger.ac.uk  
On Feb 2, 2001 this sequence version replaced gi:12329427.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em, EMBL; Sw, S.W.  
SWISSPROT; Tr, TrEMBL; Wp, WormPeP; Information on the WORMPEP  
database can be found at  
http://www.sanger.ac.uk/projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr9  
RP11-113D19 is from the library RPCT-11.1 constructed by the group  
of Pletier de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBACe3.6  
This sequence is the entire insert of clone RP11-113D19.

## FEATURES

Location/Qualifiers  
1..162064  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="9"  
/map="p21.2-22.3"

/clone="RP11-113D19"  
1225..1462  
/note="match: STS: Em:G28570"  
1457..3005  
/note="match: STS: Em:G28558"  
1464..3074  
/note="match: STS: Em:G28570"  
1846..2829  
/note="match: STS: Em:G28559"  
2441..2829  
/note="match: STS: Em:G26615"  
2444..2636  
/note="match: STS: Em:G29796"  
2624..2771  
/note="37 copies 4 mer att 60% conserved"  
4231..4346  
/note="29 copies 4 mer gata 75% conserved"  
4232..4337  
/note="53 copies 2 mer at 72% conserved"  
5200..5517  
/note="AluX repeat: matches 1..312 of consensus"  
5394..5405  
/note="Sequence confirmed by AC009445 sequenced by WBR"  
5762..6017  
/note="L1 repeat: matches 3805..5050 of consensus"  
6076..7311  
/note="L1 repeat: matches 768..975 of consensus"  
7312..7825  
/note="L1P5 repeat: matches 5270..5790 of consensus"  
7827..7866  
/note="20 copies 2 mer ta 90% conserved"  
7829..7868  
/note="10 copies 4 mer tata 90% conserved"  
7905..8129  
/note="L1MCB repeat: matches 768..975 of consensus"  
8130..8230  
/note="HY3 repeat: matches 1..101 of consensus"  
8231..8450  
/note="L1MCB repeat: matches 975..1200 of consensus"  
9409..9537  
/note="L12 repeat: matches 1267..1396 of consensus"  
9623..9754  
/note="L1M4 repeat: matches 3075..3209 of consensus"  
9755..10067  
/note="AluX repeat: matches 2..312 of consensus"  
10068..10328  
/note="L1M4 repeat: matches 2832..3075 of consensus"  
10329..10621  
/note="AluX repeat: matches 3..311 of consensus"  
10622..10739  
/note="L1M4 repeat: matches 2712..2832 of consensus"  
10801..10900  
/note="L1M4 repeat: matches 2234..2339 of consensus"  
11327..11866  
/note="L1MCB repeat: matches 1484..2015 of consensus"  
12541..12854  
/note="AluB repeat: matches 1..308 of consensus"  
13229..13385  
/note="MIR repeat: matches 54..224 of consensus"  
13781..13876  
/note="24 copies 4 mer atat 65% conserved"  
15442..15858  
/note="MER57A repeat: matches 10..433 of consensus"  
15870..16209  
/note="MER57-Internal repeat: matches 6794..7137 of  
consensus"  
16488..16783  
/note="AluB repeat: matches 1..286 of consensus"  
18013..18110  
/note="MER57-Internal repeat: matches 5953..6051 of  
consensus"  
18266..19763

```

repeat_region /note="L1PA13 repeat: matches -651. .1413 of consensus"
19765. .21781
/note="L1P repeat: matches 1580. .3600 of consensus"
repeat_region 21822. .22098
/note="Alu10 repeat: matches 1. .302 of consensus"
repeat_region 24120. .24291
/note="L1PA4 repeat: matches 5975. .6146 of consensus"
repeat_region 24292. .25119
/note="L1P2 repeat: matches 4962. .5789 of consensus"
repeat_region 25979. .26271
/note="Alu10 repeat: matches 13. .297 of consensus"
repeat_region 26472. .26525
/note="27 copies 2 mer ta 72% conserved"
misc_feature 27481. .27834
/note="match: STS: Em:G06609"
repeat_region 28014. .28089
/note="L1MA3 repeat: matches 6229. .6304 of consensus"
repeat_region 28296. .28859
/note="L2 repeat: matches 1165. .1778 of consensus"
repeat_region 28927. .29218
/note="AluSc repeat: matches 4. .293 of consensus"
repeat_region 29783. .29942
/note="L1MA6 repeat: matches 6135. .6290 of consensus"
repeat_region 29943. .30246
/note="Alu10 repeat: matches 3. .302 of consensus"
repeat_region 30547. .30739
/note="MER87 repeat: matches 168. .364 of consensus"
repeat_region 31042. .32075
/note="L1MC3 repeat: matches 5403. .6499 of consensus"
repeat_region 35918. .35987
/note="MADE1 repeat: matches 1. .79 of consensus"
repeat_region 37665. .37739
/note="L12 repeat: matches 2370. .2446 of consensus"
repeat_region 37800. .38025
/note="L1ME2 repeat: matches 5919. .6140 of consensus"
repeat_region 38026. .38377
/note="MLT1A2 repeat: matches 2. .374 of consensus"
repeat_region 38378. .39003
/note="L1ME2 repeat: matches 5274. .5919 of consensus"
repeat_region 39192. .39329
/note="L1M4 repeat: matches 4906. .5041 of consensus"
repeat_region 39357. .41264
/note="L1PA7 repeat: matches 4206. .6143 of consensus"
repeat_region 41272. .41760
/note="L1M4 repeat: matches 4387. .4884 of consensus"
repeat_region 41761. .42459
/note="L1MA8 repeat: matches 5594. .6291 of consensus"
repeat_region 43709. .44016
/note="L1M4 repeat: matches 3883. .4192 of consensus"
repeat_region 44017. .44180
/note="82 copies 2 mer aa 62% conserved"
repeat_region 44188. .44714
/note="L1 repeat: matches 4151. .4691 of consensus"
repeat_region 44743. .44912
/note="L1ME3 repeat: matches 5767. .5949 of consensus"
repeat_region 45384. .45432
/note="Charliela repeat: matches 1. .59 of consensus"
repeat_region 45442. .45755
/note="Charliela repeat: matches 2343. .2675 of consensus"
repeat_region 47465. .47528
/note="L1R29 repeat: matches 541. .602 of consensus"
repeat_region 47520. .47976
/note="MER39 repeat: matches 1. .460 of consensus"
repeat_region 48085. .48548
/note="MER39b repeat: matches 23. .579 of consensus"
repeat_region 48549. .48669
/note="L1R29 repeat: matches 5. .118 of consensus"
repeat_region 48670. .48690
/note="MER39b repeat: matches 5. .23 of consensus"
repeat_region 49642. .49786
/note="MIR repeat: matches 109. .256 of consensus"
repeat_region 49846. .52170
/note="L1Pba repeat: matches 1231. .3606 of consensus"

```

```

misc_feature 51835. .52494
/note="match: STS: Em:HSJ167F5"
repeat_region 52683. .53439
/note="L1 repeat: matches 2754. .3475 of consensus"

Query Match 1.9%: Score 20; DB 90; Length 162064;
Best Local Similarity 100.0%; Pred. NO. 8.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 734 aaaaatttcagatgacatg 753
Db 41268 AAAAGTTTCAGATGACATG 41249

```

```

RESULT 31
AC022758/c 163819 bp DNA HTG 12-MAR-2000
LOCUS Homo sapiens clone RP11-273111, WORKING DRAFT SEQUENCE, 10
DEFINITION unorderd pieces.
ACCESSION AC022758
VERSION AC022758.3 GI:7229832
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 163819)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens clone RP11-273111
JOURNAL Unpublished
2 (bases 1 to 163819)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abrahams,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Becker,T., Bede,F.,
Boguslavsky,I., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Chapel,V., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeFelice,K., Dewar,K., Domingo,M., Doyle,M., Fenebor,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Garday,S., Grant,G., Hagos,B., Heath,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Kartas,A., Klein,J.,
Landers,T., Lebecky,J., Levine,R., Liu,C., Liu,G., Locke,K.,
Macdonald,P., Margis,N., McEwan,P., McGurk,A., McKernan,K.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye-Thomann,N.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
JOURNAL Direct Submission
COMMENT Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2000 this sequence version replaced gi:6978223.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

```

----- Genome Center -----
Center: Whitehead Institute/MIT Center for Genome Research
Center code: W1BR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L5831
Center clone name: 273111
----- Summary Statistics -----
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 155915 bases at least Q40
Consensus quality: 159953 bases at least Q30
Consensus quality: 161467 bases at least Q20
Insert size: 163000; agarose-fp
Quality coverage: 4.5 in Q20 bases; agarose-fp

```

Quality coverage: 4.5 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently consists of 10 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1444 1543: contig of 1443 bp in length  
\* 1544 3909: contig of 2366 bp in length  
\* 3910 4009: gap of 100 bp  
\* 4010 9300: contig of 5291 bp in length  
\* 9301 9400: gap of 100 bp  
\* 9401 16340: contig of 6940 bp in length  
\* 16341 16440: gap of 100 bp  
\* 16441 28464: contig of 12024 bp in length  
\* 28465 28564: gap of 100 bp  
\* 28565 50316: contig of 21752 bp in length  
\* 50317 50416: gap of 100 bp  
\* 50417 74311: contig of 23895 bp in length  
\* 74312 74411: gap of 100 bp  
\* 74412 103034: contig of 28623 bp in length  
\* 103035 103134: gap of 100 bp  
\* 103135 128960: contig of 25826 bp in length  
\* 128961 129060: gap of 100 bp  
\* 129061 163819: contig of 34759 bp in length.  
Location/Qualifiers

FEATURES  
source  
1. .163819  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-273111"

misc\_feature  
1. .1443  
/note="assembly\_fragment"

misc\_feature  
1544. .3909  
/note="assembly\_fragment"

misc\_feature  
4010. .9300  
/note="assembly\_fragment"

misc\_feature  
9401. .16340  
/note="assembly\_fragment"

misc\_feature  
16441. .28464  
/note="assembly\_fragment"

misc\_feature  
28565. .50316  
/note="assembly\_fragment"

misc\_feature  
50417. .74311  
/note="assembly\_fragment"

misc\_feature  
vector\_side:left\*

misc\_feature  
74412. .103034  
/note="assembly\_fragment"

misc\_feature  
103135. .128960  
/note="assembly\_fragment"

misc\_feature  
129061. .163819  
/note="assembly\_fragment"

misc\_feature  
clone\_end:SP6  
vector\_side:right\*

BASE COUNT 52867 a 28207 c 28618 g 53220 t 907 others  
ORIGIN

Query Match 1.9%; Score 20; DB 67; Length 163819;  
Best Local Similarity 100.0%; Pred. No. 8.6;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 625 cccgattattattacatgga 644  
|||||

DB 63282 CCGATTATATATCATCGGA 63263

RESULT 32

AL157772

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AL157772 168843 bp DNA HTG 28-FEB-2001  
Homo sapiens chromosome 13 clone RP11-98F3 map q14.3-21.31, \*\*\*  
SEQUENCING IN PROGRESS \*\*\*, 8 unordered pieces.

AL157772  
AL157772.9 GI:13161566  
HTG: HTGS\_PHASE1; HTGS\_DRAFT.

human.  
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 168843)  
Burton J.

Direct Submission  
Submitted (27-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
Requests: clonerequests@sanger.ac.uk  
On Feb 28, 2001 this sequence version replaced gi:10178438.

Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk

Project Information  
Center project name: ba98P3

Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 16398 bases at least Q40  
Consensus quality: 165696 bases at least Q30  
Consensus quality: 166853 bases at least Q20  
Insert size: 168143; sum-of-contigs  
Insert size: 132867; 4.1% error; agarose-fp  
Quality coverage: 5.39x in Q20 bases; sum-of-contigs Quality  
coverage: 8.55x in Q20 bases; agarose-fp

\* NOTE: This is a 'working draft' sequence. It currently consists of 8 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 6420 6519: contig of 6419 bp in length  
\* 6520 15268: contig of 8749 bp in length  
\* 15269 15368: gap of 100 bp  
\* 15369 80649: contig of 65281 bp in length  
\* 80650 80749: gap of 100 bp  
\* 80750 157362: contig of 76613 bp in length  
\* 157363 157462: gap of 100 bp  
\* 157463 159511: contig of 2049 bp in length  
\* 159512 159611: gap of 100 bp  
\* 159612 162901: contig of 3290 bp in length  
\* 162902 163001: gap of 100 bp  
\* 163002 165635: contig of 2634 bp in length  
\* 165636 165735: gap of 100 bp  
\* 165736 168843: contig of 3108 bp in length.  
Location/Qualifiers

FEATURES  
source  
1. .168843  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="13"  
/map="q14.3-21.31"  
/clone="RP11-98F3"  
/clone\_lib="RPCT-11.1"

misc\_feature  
1. .6419  
/note="assembly\_fragment:01807  
fragment\_chain:1"

misc\_feature  
6520. .15268  
/note="assembly\_fragment:01252

```

misc_feature      fragment_chain:1"
                  15369..80649
                  /note="assembly-fragment:01624
                  fragment_chain:1"
misc_feature      80750..157362
                  /note="assembly-fragment:01810
                  fragment_chain:1"
misc_feature      157463..159511
                  /note="assembly-fragment:00214"
misc_feature      159612..162901
                  /note="assembly-fragment:00669"
misc_feature      163002..165635
                  /note="assembly-fragment:01153"
misc_feature      165736..168843
                  /note="assembly-fragment:03095"
BASE COUNT      56550 a 28798 c 28971 g 53768 t 756 others
ORIGIN

```

```

Query Match      1.9%; Score 20; DB 79; Length 168843;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      734 aaagttgcagatgactg 753
        |||
Db 163969 AAAAGTTGCAGATGACATG 163968

```

```

RESULT 33
AC021455/c      AC021455 172601 bp DNA HTG 10-JAN-2001
LOCUS           Homo sapiens clone RP11-142B17, WORKING DRAFT SEQUENCE, 18
DEFINITION      Unordered pieces.
ACCESSION       AC021455
VERSION         AC021455.4 GI:12061491
KEYWORDS        HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE          human.
ORGANISM        Homo sapiens
                Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1 (bases 1 to 172601)
AUTHORS        Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE          Unpublished
JOURNAL        2 (bases 1 to 172601)
REFERENCE       2 (bases 1 to 172601)
AUTHORS        Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
                Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,
                Boguslavsky, L., Bouckhalter, B., Brown, A., Burkett, G., Castle, A.,
                Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
                DeRubeis, K., Dewar, K., Domino, M., Doyle, M., Feneclor, J.,
                Fertelino, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,
                Gardyna, S., Grant, G., Hagos, B., Heatford, A., Horton, L.,
                Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
                Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K.,
                McDonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
                McPheters, R., Meltrin, J., Menes, L., Morrow, J., Naylor, J.,
                Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,
                Pierre, N., Pisan, C., Pollara, Y., Raymond, C., Riley, R., Rothman, D.,
                Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
                Stojanovic, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J.,
                Tittel, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
                Zimmer, A. and Zody, M.

```

```

TITLE          Direct Submission
JOURNAL        Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
                Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT        On Jan 10, 2001 this sequence version replaced gi:1417782.
                All repeats were identified using RepeatMasker:
                http://ftp.genome.washington.edu/RM/RepeatMasker.html
                ----- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu

```

```

Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5173
Center clone name: 142_B-17
----- Summary Statistics
Sequencing vector: M13; M77815; 59% of reads
Sequencing vector: Plasmid; n/a; 41% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 16026 bases at least Q40
Consensus quality: 168757 bases at least Q20
Consensus quality: 169806 bases at least Q20
Insert size: 17000; agarose-fp
Insert size: 170901; sum-of-ctrls
Quality coverage: 7.3 in Q20 bases; agarose-fp
Quality coverage: 7.2 in Q20 b.
NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 34725: contig of 34725 bp in length
34726 34825: gap of 100 bp
34826 348305: contig of 480 bp in length
35306 35405: gap of 100 bp
35406 36203: contig of 798 bp in length
36204 36303: gap of 100 bp
36304 38839: contig of 2536 bp in length
38840 38939: gap of 100 bp
38940 40077: contig of 1138 bp in length
40078 40177: gap of 100 bp
40178 44051: contig of 3874 bp in length
44052 44151: gap of 100 bp
44152 46550: contig of 2399 bp in length
46551 46650: gap of 100 bp
46651 52179: contig of 5529 bp in length
52180 52279: gap of 100 bp
52280 58781: contig of 6502 bp in length
58782 58881: gap of 100 bp
58882 66968: contig of 8087 bp in length
66969 67068: gap of 100 bp
67069 77112: contig of 10044 bp in length
77113 77212: gap of 100 bp
77213 89138: contig of 11926 bp in length
89139 89238: gap of 100 bp
89239 98499: contig of 9261 bp in length
98500 98599: gap of 100 bp
98600 108114: contig of 9515 bp in length
108115 108214: gap of 100 bp
108215 117592: contig of 9378 bp in length
117593 117692: gap of 100 bp
117693 132829: contig of 15137 bp in length
132830 132929: gap of 100 bp
132930 147541: contig of 14612 bp in length
147542 147641: gap of 100 bp
147642 172601: contig of 24960 bp in length.
FEATURES
source
1..172601
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-142B17"
/clone_11b="RP11 Human Male BAC"
1..34725
/note="assembly-fragment"
clone_end:SP6
vector_side:left"
misc_feature      34826..35305
                  /note="assembly-fragment"
misc_feature      35406..36203
                  /note="assembly-fragment"

```





```

misc_feature      /note="assembly_fragment"
                  12425..20299
                  /note="assembly_fragment"
misc_feature      20400..41002
                  /note="assembly_fragment"
                  clone_end:17
                  vector_side:right"
misc_feature      41103..172758
                  /note="assembly_fragment"
                  clone_end:SP6
                  vector_side:right"
BASE COUNT      55711 a 30643 c 31348 g 54356 t 700 others
ORIGIN

```

```

Query Match      1.9% Score 20; DB 67; Length 172758;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 547 ttcttctgaactgcaagt 566
Db 55356 TTCTCTTGACATGCAAGT 55337

```

```

RESULT 35
AC020624/c      AC020624 180977 bp DNA HTG 10-NCV-2000
LOCUS          Homo sapiens chromosome 3 clone RP11-286C5, WORKING DRAFT SEQUENCE,
DEFINITION     15 unordered pieces.
AC020624
VERSION        AC020624.11 GI:11128296
KEYWORDS       HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE         human.
ORGANISM       Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 180977)
AUTHORS       Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
                Alstrooks,S.L., Amaralunga,H.C., Are,J.R., Banks,T., Barbara,J.,
                Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
                Bowle,S., Brieval,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
                Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carion,T.F.,
                Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
                Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
                Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
                Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,C.,
                Demn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
                Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
                Ellhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
                Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
                Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
                Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
                Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
                Hollins,B., Honsl,F., Howard,S., Huber,J., Hulik,S., Hume,J.,
                Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolyet,S.,
                Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,I.,
                Kovar,C., Kratovic,J., Kuresh,A., Landry,N., Leal,B., Lewis,L.C.,
                Lewis,J., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,X., Liu,W.,
                Louised,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
                Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
                Martinez,E., Massey,E., Maswney,E., McLeod,M.P., Meador,M.,
                Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
                Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
                Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nokenko,S.,
                Ogutu,M., Okunonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
                Peetey,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
                Puelles,M., Ren,Y., Rives,M., Rojas,A., Rojudoakan,I., Rolfe,M.,
                Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shoshari,N.,
                Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,
                Stone,H., Sutton,A., Svatek,A., Tabors,P., Tamaria,A., Tamerisa,K.,
                Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
                Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
                Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,

```

```

TITLE           Wallington,S., Williams,G., Williamson,A., Mieczky,R., Wooden,S.,
JOURNAL         Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
                and Gibbs,R.
DIRECT SUBMISSION
UNPUBLISHED
REFERENCE       2 (bases 1 to 180977)
AUTHORS        Worley,K.C.
TITLE          Direct Submission
JOURNAL        Submitted (07-JAN-2000) Human Genome Sequencing Center, Department
                of Molecular and Human Genetics, Baylor College of Medicine, One
                Baylor Plaza, Houston, TX 77030, USA
                On Nov 9, 2000 this sequence version replaced gi:11067223.
COMMENT         ----- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web Site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
                ----- Project Information
                Center project name: HMXN
                Center clone name: RP11-286C5
                ----- Summary Statistics
                Sequencing vector: M13: 108821
                Chemistry: Dye-primer Bodypy: 9% of reads
                Chemistry: Dye-terminator Big Dye: 91% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 159337 bases at least Q40
                Consensus quality: 170570 bases at least Q30
                Consensus quality: 174900 bases at least Q20
                Estimated insert size: 176988; sum-of-ctggs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 4x in Q20 bases; sum-of-ctggs estimation

```

```

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 35066: contig of 35066 bp in length
35067 35166: gap of unknown length
35167 64945: contig of 29779 bp in length
64946 65045: gap of unknown length
65046 89710: contig of 24665 bp in length
89711 89810: gap of unknown length
89811 109592: contig of 19782 bp in length
109593 109692: gap of unknown length
109693 122681: contig of 12989 bp in length
122682 122781: gap of unknown length
122782 133919: contig of 11138 bp in length
133920 134020: gap of unknown length
134021 142391: contig of 8372 bp in length
142392 142491: gap of unknown length
142492 150079: contig of 7588 bp in length
150080 150179: gap of unknown length
150180 158317: contig of 8138 bp in length
158318 158317: gap of unknown length
158318 162652: contig of 7845 bp in length
162653 162652: gap of unknown length
162653 169778: contig of 3416 bp in length
169779 169878: gap of unknown length
169879 174812: contig of 4294 bp in length
174813 174272: gap of unknown length
174273 177641: contig of 3369 bp in length
177642 177742: gap of unknown length
177743 179542: contig of 1801 bp in length
179543 179643: gap of unknown length
179644 180977: contig of 1335 bp in length.
Location/Qualifiers
1. 180977

```

FEATURES  
Source

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/clone="RP11-286G5"  
BASE COUNT 54183 a 35366 c 35783 g 54181 t 1464 others  
ORIGIN

Query Match 1.9%; Score 20; DB 66; Length 180977;  
Best Local Similarity 100.0%; Pred. No. 8.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 625 cctgattattacatgga 644  
|||||  
db 33003 CCGATTATATCTCTGGA 32984

RESULT 36  
AC036195 185621 bp DNA HTG 26-MAY-2000  
LOCUS  
DEFINITION Homo sapiens chromosome 4 clone RP11-460L16 map 4, WORKING DRAFT  
SEQUENCE 33 unordered pieces.  
AC036195  
AC036195.2 GI:8082698  
VERSION HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS Birren, B., Linton, L., Nussbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F.,  
Bouckling, L., Bouckling, B., Brown, A., Burkett, G.,  
Campomaior, A., Castle, A., Choquet, Y., Colangelo, M., Collins, S.,  
Collamore, A., Cooke, P., DeArle, K., Dewar, K., Diaz, J.S.,  
Dodge, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,  
Galagan, S., Gardina, S., Ginde, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
Howard, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
Klein, J., Larocque, K., Lamas, R., Landers, T., Lebecky, J.,  
Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,  
McCarthy, M., McEwan, P., McGuck, A., McKernan, K., McPherson, R.,  
Meldrum, J., Meneu, L., Mihova, T., Miranda, C., Mlenka, V., Morrow, J.,  
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,  
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Testa, S., Theodore, J., Tittel, A., Travers, M., Triggillo, J.,  
Vassiliev, H., Viel, R., Vo, A., Willson, B., Wu, X., Wyman, D., Ye, W.J.,  
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
COMMENT  
Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 25, 2000 this sequence version replaced gi:7523864.  
All repeats were identified using RepeatMasker.  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
Project Information  
Center Project name: 19288  
Center Clone name: 460.L\_16  
Sequencing Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731  
Consensus quality: 166388 bases at least Q40  
Consensus quality: 176238 bases at least Q30  
Consensus quality: 180267 bases at least Q20  
Insert size: 194000; agarose-gel  
Insert size: 182421; sum-of-contigs  
Quality coverage: 3.3 in Q20 bases; agarose-gel  
Quality coverage: 3.5 in Q20 bases; sum-of-contigs  
NOTE: This is a 'working draft' sequence. It currently  
consists of 33 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.  
1 1092: contig of 1092 bp in length  
\* 1093 1197: gap of 100 bp  
\* 1193 2761: contig of 1569 bp in length  
\* 2762 2861: gap of 100 bp  
\* 2862 4784: contig of 1923 bp in length  
\* 4785 4884: gap of 100 bp  
\* 4885 7240: contig of 2356 bp in length  
\* 7241 7340: gap of 100 bp  
\* 7341 8766: contig of 1426 bp in length  
\* 8767 8866: gap of 100 bp  
\* 8867 10759: contig of 1893 bp in length  
\* 10760 10859: gap of 100 bp  
\* 10860 13875: contig of 3016 bp in length  
\* 13876 13975: gap of 100 bp  
\* 13976 15726: contig of 1751 bp in length  
\* 15727 15826: gap of 100 bp  
\* 15827 18056: contig of 2230 bp in length  
\* 18057 18156: gap of 100 bp  
\* 18157 19481: contig of 1325 bp in length  
\* 19482 19581: gap of 100 bp  
\* 19582 20346: contig of 765 bp in length  
\* 20347 20446: gap of 100 bp  
\* 20447 22218: contig of 1772 bp in length  
\* 22219 22318: gap of 100 bp  
\* 22319 25876: contig of 3558 bp in length  
\* 25877 25976: gap of 100 bp  
\* 25977 30791: contig of 4815 bp in length  
\* 30792 30891: gap of 100 bp  
\* 30892 34576: contig of 3685 bp in length  
\* 34577 34676: gap of 100 bp  
\* 34677 38131: contig of 3455 bp in length  
\* 38132 38231: gap of 100 bp  
\* 38232 41823: contig of 3592 bp in length  
\* 41824 41923: gap of 100 bp  
\* 41924 46113: contig of 4190 bp in length  
\* 46114 46213: gap of 100 bp  
\* 46214 52312: contig of 6099 bp in length  
\* 52313 52412: gap of 100 bp  
\* 52413 56468: contig of 4056 bp in length  
\* 56469 56568: gap of 100 bp  
\* 56569 61405: contig of 4837 bp in length  
\* 61406 61505: gap of 100 bp  
\* 61506 67256: contig of 5751 bp in length  
\* 67257 67356: gap of 100 bp  
\* 67357 73099: contig of 5743 bp in length  
\* 73100 73199: gap of 100 bp  
\* 73200 80870: contig of 7671 bp in length  
\* 80871 80970: gap of 100 bp  
\* 80971 88670: contig of 7700 bp in length  
\* 88671 88770: gap of 100 bp  
\* 88771 96686: contig of 7916 bp in length  
\* 96687 96786: gap of 100 bp  
\* 96787 106092: contig of 9306 bp in length  
\* 106093 106192: gap of 100 bp  
\* 106193 114414: contig of 8222 bp in length  
\* 114415 114514: gap of 100 bp

```

FEATURES
source
* 114515 127090: contig of 12576 bp in length
* 127091 127190: gap of 100 bp
* 127191 137141: contig of 9951 bp in length
* 137142 137241: gap of 100 bp
* 137242 149169: contig of 11928 bp in length
* 149170 149289: gap of 100 bp
* 149290 166103: contig of 16834 bp in length
* 166104 166203: gap of 100 bp
* 166204 185621: contig of 19418 bp in length.
Location/Qualifiers
1. 185621
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-460L16"
/clone_id="RP11 Human Male BAC"
1. 1092
/misc_feature
/feature="assembly_fragment"
1193. 2761
/misc_feature
/feature="assembly_fragment"
2862. 4784
/misc_feature
/feature="assembly_fragment"
4885. 7240
/misc_feature
/feature="assembly_fragment"
7341. 8766
/misc_feature
/feature="assembly_fragment"
8867. 10759
/misc_feature
/feature="assembly_fragment"
10860. 13875
/misc_feature
/feature="assembly_fragment"
13976. 15726
/misc_feature
/feature="assembly_fragment"
15827. 18056
/misc_feature
/feature="assembly_fragment"
18157. 19481
/misc_feature
/feature="assembly_fragment"
19582. 20346
/misc_feature
/feature="assembly_fragment"
clone_end:17
vector_side:left"
20447. 22218
/misc_feature
/feature="assembly_fragment"
22319. 25876
/misc_feature
/feature="assembly_fragment"
25977. 30791
/misc_feature
/feature="assembly_fragment"
30892. 34576
/misc_feature
/feature="assembly_fragment"
34677. 38131
/misc_feature
/feature="assembly_fragment"
38232. 41823
/misc_feature
/feature="assembly_fragment"
41924. 46113
/misc_feature
/feature="assembly_fragment"
46214. 52312
/misc_feature
/feature="assembly_fragment"
52413. 56468
/misc_feature
/feature="assembly_fragment"
56569. 61405
/misc_feature
/feature="assembly_fragment"
61506. 67256
/misc_feature
/feature="assembly_fragment"
67357. 73099
/misc_feature
/feature="assembly_fragment"
clone_end:576
vector_side:right"
73200. 80870
/misc_feature
/feature="assembly_fragment"
80971. 88670
/misc_feature
/feature="assembly_fragment"
88771. 96686
/misc_feature
/feature="assembly_fragment"

```

```

Query Match 1.98; Score 20; DB 71; Length 185621;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 547 ttcttctgactctgaagt 566
|||||
DB 44499 TTCTTCTGACACTCAAGT 44518

```

```

RESULT 37
AL589844/C
LOCUS
DEFINITION
Homo sapiens chromosome 9 clone RP11-436F21, *** SEQUENCING IN
PROGRESS ***, 2 unordered pieces.
ACCESSION
AL589844 AC020712
VERSION
AL589844.2 GI:13561354
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 187892)
REFERENCE
1 (bases 1 to 187892)
AUTHORS
Plumb, B.
JOURNAL
Direct Submission
Submitted (11-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Apr 8, 2001 this sequence version replaced gi:13374066.
Draft Sequence Produced by Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
http://genome.wustl.edu/gsc/index.shtml
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BA436F21
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M7815; 19% of reads
Sequencing vector: Plasmid; L08752; 80% of reads
Chemistry: Dye-terminator Big Dye; 82% of reads
Chemistry: Dye-primer-amerisham; 17% of reads
Consensus quality: 187443 bases at least Q40
Consensus quality: 187619 bases at least Q30
Consensus quality: 187719 bases at least Q20
Insert size: 187792; sum-of-contigs
Insert size: 175585; 3.0% error; agarose-1p
Quality coverage: 15.52x in Q20 bases; sum-of-contigs quality
coverage: 16.60x in Q20 bases; agarose-1p
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 151810: contig of 151810 bp in length
* 151811 151910: gap of 100 bp
* 151911 187892: contig of 35982 bp in length.
Location/Qualifiers
1. 187892
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-436F21"
/clone_id="RP11-436F21"
1. 151810
/misc_feature
/feature="assembly_fragment"

```

/note="assembly fragment:00911  
fragment chain:1  
clone\_end:T7  
vector\_side:left"  
151911..187892  
/note="assembly fragment:04440  
fragment chain:1  
clone\_end:SP6  
vector\_side:right"

BASE COUNT 61201 a 33417 c 32695 g 60479 t 100 others  
ORIGIN

Query Match 1.9%; Score 20; DB 82; Length 187892;  
Best Local Similarity 100.0%; Pired. No. 8.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 625 cctgattatattactgga 644  
|||||  
Db 33347 CCTGATTATATCTACTGCA 33328

RESULT 38  
CNS01DX4 190515 bp DNA HTG 09-OCT-2000  
LOCUS Homo sapiens chromosome 14 clone R-840119, \*\*\* SEQUENCING IN  
DEFINITION PROGRESS \*\*\* in ordered pieces.  
ACCESSION AL139022  
VERSION AL139022.3 GI:10798506  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 190515)  
Genoscope.  
Direct Submission  
Submitted (09-OCT-2000) Genoscope - Centre National de Sequencage :  
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
On Oct 11, 2000 this sequence version replaced gi:9212124.  
----- Genome Center  
Center: Genoscope / Centre National de Sequencage  
Center code: GS  
Web site: http://www.genoscope.cns.fr/  
Contact: SeqRef@genoscope.cns.fr

IMPORTANT: This sequence is unfinished and does not necessarily  
represent the correct sequence. Work on the sequence is in progress  
and the release of this data is based on the understanding that the  
sequence may change as work continue. The sequence may be  
contaminated with foreign sequence from E.coli, yeast, vector,  
phage, etc. . even if efforts are made to eliminate these  
contaminating sequences. The following BAC sequence is oriented  
from the T7 to the SP6 end.  
Upstream BAC (overlapping the T7 end) : C-2509616 (AC=AL355076)  
Downstream BAC (overlapping the SP6 end) : C-2300N4  
Summary Statistics  
Assembly Program: Phrap; version 2.0  
Quality coverage: 6.82x in Q20 bases; sum-of-contigs

-----  
Overall quality chart :  
Range : bases  
0 : 126  
10 - 9 : 67  
10 - 19 : 505  
20 - 29 : 1460  
30 - 39 : 5233  
40 - 49 : 14778  
50 - 59 : 17050  
60 - 69 : 11239  
70 - 79 : 18989

80 - 89 : 47427  
90 - 99 : 73641  
-----  
Percentage of bases with a quality value >= 40 : 96 %.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
Location/Qualifiers  
1. .190515

FEATURES  
source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="14"  
/clone="R-840119"  
/clone\_1fb="pPC1-11"  
22164..22308  
/note="matching EMBL:H99639  
Rdb:RH68943  
dbSTS:STS48849  
Identified using the e-PCR software (G. Schuler)"  
30402..30625  
/note="matching EMBL:R42375  
Rdb:RH69913  
dbSTS:STS49790  
Identified using the e-PCR software (G. Schuler)"  
30773..30950  
/note="matching EMBL:AA053315  
Rdb:RH67850  
dbSTS:STS47777  
Identified using the e-PCR software (G. Schuler)"  
43553..43740  
/note="matching EMBL:Z39082  
Rdb:RH25206  
dbSTS:STS25011  
Identified using the e-PCR software (G. Schuler)"  
44713..44833  
/note="matching EMBL:L00635  
Rdb:RH53755  
dbSTS:STS12411  
Identified using the e-PCR software (G. Schuler)"  
44713..44853  
/note="matching EMBL:L00635  
Rdb:RH53870  
dbSTS:STS42913  
Identified using the e-PCR software (G. Schuler)"  
47223..47412  
/note="matching EMBL:AA165116  
Rdb:RH48540  
dbSTS:STS41599  
Identified using the e-PCR software (G. Schuler)"  
128323..128438  
/note="matching EMBL:G07650  
Rdb:RH34434  
dbSTS:STS6376  
Identified using the e-PCR software (G. Schuler)"  
164136..164258  
/note="matching EMBL:AA160869  
Rdb:RH48489  
dbSTS:STS41548  
Identified using the e-PCR software (G. Schuler)"  
169171..169324  
/note="matching EMBL:W56308  
Rdb:RH47847  
dbSTS:STS40906  
Identified using the e-PCR software (G. Schuler)"  
171253..171385  
/note="matching EMBL:Z39818  
Rdb:RH10964  
dbSTS:STS1017  
Identified using the e-PCR software (G. Schuler)"  
173565..173714  
/note="matching EMBL:R10900  
Rdb:RH53901

## STS

dbSTS:STS22552  
 Identified using the e-PCR software (G. Schuler)\*  
 17384..174013  
 /note="matching EXML:G27797  
 RhdB:RH32719  
 dbSTS:STS4675  
 Identified using the e-PCR software (G. Schuler)\*  
 52743 a 43156 c 42563 g 52044 t 9 others

## BASE COUNT

Query Match 1.9% Score 20: DB 83: Length 190515;  
 Best Local Similarity 100.0% Pred. No. 8.6;  
 Matches 20: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 Qy 52 gagttagatataataac 71  
 Db 35077 GAGTTAGTATATTAATC 35096

## RESULT 39

AC007448 190802 bp DNA HTG 03-MAR-2001  
 LOCUS Homo sapiens chromosome 17 clone RP11-401F2 map 17, WORKING DRAFT  
 DEFINITION  
 AC007448  
 AC007448.3 GI:13184199  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

Consensus quality: 180653 bases at least Q40  
 Consensus quality: 185418 bases at least Q30  
 Consensus quality: 187087 bases at least Q20  
 Insert size: 186000; agarose-fp  
 Insert size: 188202; sum-of-contigs  
 Quality coverage: 6.6 in Q20 bases.  
 NOTE: This is a 'working draft' sequence. It currently  
 consists of 27 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1 30976: contig of 30976 bp in length  
 30977 31076: gap of 100 bp  
 31077 31957: contig of 881 bp in length  
 31958 32057: gap of 100 bp  
 32058 32844: contig of 787 bp in length  
 32845 32944: gap of 100 bp  
 32945 33411: contig of 467 bp in length  
 33412 33511: gap of 100 bp  
 33512 34592: contig of 1081 bp in length  
 34593 34692: gap of 100 bp  
 34693 35847: contig of 1155 bp in length  
 35848 35947: gap of 100 bp  
 35948 37177: contig of 1230 bp in length  
 37178 37277: gap of 100 bp  
 37278 37948: contig of 671 bp in length  
 37949 38048: gap of 100 bp  
 38049 39288: contig of 1240 bp in length  
 39289 39388: gap of 100 bp  
 39389 41514: contig of 2126 bp in length  
 41515 41614: gap of 100 bp  
 41615 44038: contig of 2424 bp in length  
 44039 44138: gap of 100 bp  
 44139 45853: contig of 1715 bp in length  
 45854 45953: gap of 100 bp  
 45954 48151: contig of 2198 bp in length  
 48152 48251: gap of 100 bp  
 48252 51380: contig of 3129 bp in length  
 51381 51480: gap of 100 bp  
 51481 55357: contig of 3877 bp in length  
 55358 55457: gap of 100 bp  
 55458 58404: contig of 2947 bp in length  
 58405 58504: gap of 100 bp  
 58505 65466: contig of 6662 bp in length  
 65467 65566: gap of 100 bp  
 65567 65668: contig of 4302 bp in length  
 65669 69688: gap of 100 bp  
 69689 76614: contig of 6646 bp in length  
 76615 76714: gap of 100 bp  
 76715 87092: contig of 10378 bp in length  
 87093 87192: gap of 100 bp  
 87193 94413: contig of 7221 bp in length  
 94414 94513: gap of 100 bp  
 94514 103613: contig of 9100 bp in length  
 103614 103713: gap of 100 bp  
 103714 112413: contig of 8700 bp in length  
 112414 112513: gap of 100 bp  
 112514 121947: contig of 9434 bp in length  
 121948 122047: gap of 100 bp  
 122048 136209: contig of 14162 bp in length  
 136210 136309: gap of 100 bp  
 136310 157564: contig of 21255 bp in length  
 157565 157664: gap of 100 bp  
 157665 190802: contig of 33138 bp in length.

## FEATURES

## source

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIDR  
 Web site: <http://www.seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 Project Information  
 Center project name: L747  
 Center clone name: 401\_F\_2  
 Summary Statistics  
 Sequencing vector: M13: M77815: 58% of reads  
 Sequencing vector: plasmid: n/a: 42% of reads  
 Chemistry: Dye-terminator Big Dye: 38% of reads  
 Assembly program: Phrap: version 0.960731

```

misc_feature      /clone="Rp11-401F2"
                  /clone_lib="RP11-11 Human Male BAC"
                  /note="assembly-fragment"
                  1. .30976
                  vector_end:sp6
                  vector_start:left"
misc_feature      31077. .31957
                  /note="assembly-fragment"
misc_feature      32058. .32844
                  /note="assembly-fragment"
misc_feature      32945. .33411
                  /note="assembly-fragment"
misc_feature      33512. .34592
                  /note="assembly-fragment"
misc_feature      34693. .35847
                  /note="assembly-fragment"
misc_feature      35948. .37177
                  /note="assembly-fragment"
misc_feature      37278. .37948
                  /note="assembly-fragment"
misc_feature      38049. .39288
                  /note="assembly-fragment"
misc_feature      39389. .41514
                  /note="assembly-fragment"
misc_feature      41615. .44038
                  /note="assembly-fragment"
misc_feature      44139. .45853
                  /note="assembly-fragment"
misc_feature      45954. .48151
                  /note="assembly-fragment"
misc_feature      48252. .51380
                  /note="assembly-fragment"
misc_feature      51481. .53357
                  /note="assembly-fragment"
misc_feature      55458. .58404
                  /note="assembly-fragment"
misc_feature      58505. .65466
                  /note="assembly-fragment"
misc_feature      65567. .69868
                  /note="assembly-fragment"
misc_feature      69969. .76614
                  /note="assembly-fragment"
misc_feature      76715. .87092
                  /note="assembly-fragment"
misc_feature      87193. .94413
                  /note="assembly-fragment"
misc_feature      94514. .103613
                  /note="assembly-fragment"
misc_feature      103714. .112413
                  /note="assembly-fragment"
misc_feature      112514. .121947
                  /note="assembly-fragment"
misc_feature      122048. .136209
                  /note="assembly-fragment"
misc_feature      136310. .157564
                  /note="assembly-fragment"
misc_feature      157665. .190802
                  /note="assembly-fragment"
misc_feature      /note="assembly-fragment"
                  clone_end:17
                  vector_start:right"
BASE COUNT      53832 a 40797 c 38520 g 55048 t 2605 others
ORIGIN

```

Query Match 1.9%; Score 20; DB 60; Length 190802;  
 Best Local Similarity 100.0%; Pred. No. 8.6;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 480 agaaactgttcactgtt 499  
 ||||||||||||||||  
 Db 73481 AGAAACTGTCTCAGCTT 73462

```

RESULT 40
AC079959
LOCUS      AC079959 219888 bp DNA HTG 12-DEC-2000
DEFINITION Mus musculus clone Rp23-327121, WORKING DRAFT SEQUENCE, 4 unordered
            pieces.
ACCESSION  AC079959
VERSION    AC079959.9 GI:11641352
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 219888)
            Loh, P., Qi, S., Ford, B., and Roe, B. A.
            Mus musculus BAC Clone Rp23-327121
            Unpublished
REFERENCE  2 (bases 1 to 219888)
            Loh, P., Qi, S., Ford, B., and Roe, B. A.
            Direct Submission
            Submitted (20-SEP-2000) Department of Chemistry And Biochemistry,
            The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
            On Dec 12, 2000 this sequence version replaced gi:11321791.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 1 23291: contig of 23291 bp in length
            * 23292 23391: gap of unknown length
            * 23392 66837: contig of 43446 bp in length
            * 66838 66937: gap of unknown length
            * 66938 111800: contig of 44863 bp in length
            * 111801 111900: gap of unknown length
            * 111901 219888: contig of 107988 bp in length.
            Location/Qualifiers
                source
                1. .219888
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /clone="Rp23-327121"
                /clone_lib="Rp23"
BASE COUNT    63796 a 45318 c 45391 g 64873 t 310 others
ORIGIN

```

Query Match 1.9%; Score 20; DB 76; Length 219888;  
 Best Local Similarity 100.0%; Pred. No. 8.6;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 ttcttgaggaactcgtt 369  
 ||||||||||||||||  
 Db 77589 TTCTTGAGGAACCTCCTT 77608

```

RESULT 41
AL162424
LOCUS      AL162424 223020 bp DNA PRI 01-NOV-2000
DEFINITION Human DNA sequence from clone RP11-498E2 on chromosome 9, complete
            sequence.
ACCESSION  AL162424
VERSION    AL162424.20 GI:11042769
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 223020)
            Lloyd, D.
            Direct Submission
            Submitted (31-OCT-2000) Sanger Centre, Hinxton, Cambridgeshire,

```

CB10 15A, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Oct 30, 2000 this sequence version replaced gi:10799301.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMREP; Information on the WORMREP database can be found at

[http://www.sanger.ac.uk/projects/C\\_elegans/wormrep](http://www.sanger.ac.uk/projects/C_elegans/wormrep) This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr9>

RP11-498E2 is from the library RPC11.2 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>

VECTOR: pBAC3.6

This sequence is the entire insert of clone RP11-498E2. The true left end of clone RP11-563N7 is at 124365 in this sequence. The true left end of clone RP11-542K23 is at 222308 in this sequence. The true right end of clone RP11-456D21 is at 99792 in this sequence.

#### FEATURES

##### Source

1. 223020  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="9"  
/clone="RP11-498E2"  
/clone\_1lb="RPC1-11.2"  
513. 839  
repeat\_region  
/note="AluYb repeat: matches 1. .311 of consensus"  
865. 1091  
repeat\_region  
/note="L2 repeat: matches 2521. .2749 of consensus"  
1143. 1505  
repeat\_region  
/note="TMR1B repeat: matches 1. .364 of consensus"  
2293. 2589  
repeat\_region  
/note="AluSc repeat: matches 1. .296 of consensus"  
2590. 2760  
repeat\_region  
/note="AluSg/x repeat: matches 129. .303 of consensus"  
2835. 2929  
repeat\_region  
/note="MIR repeat: matches 49. .127 of consensus"  
2930. 3407  
repeat\_region  
/note="MTR1D repeat: matches 1. .505 of consensus"  
3408. 3495  
repeat\_region  
/note="MIR repeat: matches 127. .262 of consensus"  
3731. 3828  
repeat\_region  
/note="MIR repeat: matches 50. .146 of consensus"  
3969. 4007  
repeat\_region  
/note="MIR repeat: matches 224. .262 of consensus"  
4008. 4527  
repeat\_region  
/note="MTR2CB repeat: matches 1. .501 of consensus"  
4528. 4668  
repeat\_region  
/note="MIR repeat: matches 86. .224 of consensus"  
5125. 5252  
repeat\_region  
/note="MIR repeat: matches 20. .147 of consensus"  
5253. 5557  
repeat\_region  
/note="AluSx repeat: matches 1. .307 of consensus"  
5558. 5673  
repeat\_region  
/note="MIR repeat: matches 147. .262 of consensus"  
6071. 6214  
repeat\_region  
/note="L1PA7 repeat: matches 5998. .6143 of consensus"  
6215. 6250  
repeat\_region

repeat\_region  
/note="18 copies 2 mer aa 80% conserved"  
6470. 6561  
repeat\_region  
/note="MIR repeat: matches 20. .109 of consensus"  
6565. 6776  
repeat\_region  
/note="L2 repeat: matches 1439. .1659 of consensus"  
7006. 7060  
repeat\_region  
/note="L2 repeat: matches 2695. .2750 of consensus"  
7181. 7496  
repeat\_region  
/note="AluIo repeat: matches 1. .310 of consensus"  
7600. 8001  
repeat\_region  
/note="L2 repeat: matches 2243. .2699 of consensus"  
9047. 9195  
repeat\_region  
/note="MIR repeat: matches 70. .250 of consensus"  
10132. 10561  
repeat\_region  
/note="L2 repeat: matches 2327. .2750 of consensus"  
10816. 10880  
repeat\_region  
/note="MER4-internal repeat: matches 1095. .1160 of consensus"  
11069. 11375  
repeat\_region  
/note="AluY repeat: matches 1. .303 of consensus"  
11724. 11800  
repeat\_region  
/note="L2 repeat: matches 2555. .2631 of consensus"  
11834. 12140  
repeat\_region  
/note="AluSg repeat: matches 1. .307 of consensus"  
12669. 12944  
repeat\_region  
/note="AluSg repeat: matches 25. .300 of consensus"  
13315. 13440  
repeat\_region  
/note="F1AM\_A repeat: matches 1. .126 of consensus"  
13443. 13742  
repeat\_region  
/note="AluSx repeat: matches 1. .301 of consensus"  
13770. 14108  
repeat\_region  
/note="HERVL repeat: matches 5112. .5443 of consensus"  
14195. 14634  
repeat\_region  
/note="L2 repeat: matches 2096. .2567 of consensus"  
15311. 15821  
repeat\_region  
/note="AluSx repeat: matches 1. .311 of consensus"  
15933. 16184  
repeat\_region  
/note="L2 repeat: matches 2481. .2741 of consensus"  
16879. 17130  
repeat\_region  
/note="MIR repeat: matches 7. .258 of consensus"  
17674. 17704  
repeat\_region  
/note="MIR repeat: matches 205. .233 of consensus"  
17705. 18003  
repeat\_region  
/note="AluSx repeat: matches 1. .299 of consensus"  
18004. 18172  
repeat\_region  
/note="MIR repeat: matches 20. .205 of consensus"  
18242. 18322  
repeat\_region  
/note="L2 repeat: matches 2624. .2709 of consensus"  
19353. 19480  
repeat\_region  
/note="MIR repeat: matches 35. .174 of consensus"  
19397. 19832  
repeat\_region  
/note="match: STS: Em:HSJ42H8"  
19675. 19774  
repeat\_region  
/note="L1MC/D repeat: matches 5719. .5818 of consensus"  
19849. 20087  
repeat\_region  
/note="MIR repeat: matches 31. .262 of consensus"  
20290. 20458  
repeat\_region  
/note="AluY repeat: matches 135. .301 of consensus"  
20459. 20761  
repeat\_region  
/note="AluY repeat: matches 1. .299 of consensus"  
20762. 20895  
repeat\_region  
/note="AluY repeat: matches 1. .135 of consensus"  
20959. 21055  
repeat\_region  
/note="MIR repeat: matches 47. .146 of consensus"  
21443. 21612  
repeat\_region  
/note="MIR repeat: matches 1. .174 of consensus"  
22154. 22412  
repeat\_region  
/note="MIR repeat: matches 16. .260 of consensus"  
23468. 23637  
repeat\_region  
/note="MIR repeat: matches 12. .194 of consensus"  
24151. 24449  
repeat\_region  
/note="AluSx repeat: matches 3. .301 of consensus"  
24553. 24839  
repeat\_region

repeat\_region /note="AluJo repeat: matches 1. .288 of consensus"  
24840. .25141  
/note="AluY repeat: matches 1. .302 of consensus"  
25320. .25402  
repeat\_region /note="MIR repeat: matches 165. .247 of consensus"  
26500. .26698  
repeat\_region /note="MIR repeat: matches 30. .148 of consensus"  
27316. .27436  
repeat\_region /note="MIR repeat: matches 2. .134 of consensus"  
27526. .27691  
repeat\_region /note="MIR repeat: matches 2. .211 of consensus"  
28542. .28834  
repeat\_region /note="AluSc repeat: matches 2. .294 of consensus"  
29490. .29778  
repeat\_region /note="L2 repeat: matches 2432. .2750 of consensus"  
29858. .29973  
repeat\_region /note="L2 repeat: matches 2609. .2714 of consensus"  
29974. .30270  
repeat\_region /note="AluX repeat: matches 1. .296 of consensus"  
30271. .30330  
repeat\_region /note="L2 repeat: matches 2547. .2609 of consensus"  
30444. .31116  
repeat\_region /note="L1MC3 repeat: matches 7005. .7738 of consensus"  
31195. .31494  
repeat\_region /note="AluYb repeat: matches 1. .311 of consensus"  
31572. .31865  
repeat\_region /note="AluSc repeat: matches 2. .296 of consensus"  
32173. .32243  
repeat\_region /note="L2 repeat: matches 2624. .2750 of consensus"  
33927. .34124  
repeat\_region /note="L1MC3 repeat: matches 5590. .5904 of consensus"  
34396. .34563  
repeat\_region /note="MIR repeat: matches 20. .188 of consensus"  
35338. .35502  
repeat\_region /note="MIR repeat: matches 7. .186 of consensus"  
35560. .35648  
repeat\_region /note="MIR repeat: matches 95. .200 of consensus"  
35992. .36287  
repeat\_region /note="AluX repeat: matches 1. .296 of consensus"  
36301. .36410

Query Match 1.9%; Score 20; DB 90; Length 223020;  
Best Local Similarity 100.0%; Pred. No. 8.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 966 tatcacccttacaattt 985  
|||||  
DB 80888 TATCACCCTCTACAAATTT 80907

RESULT 42  
CNS074C0 855 bp DNA STS 11-JAN-2001  
LOCUS clone BA0AB026G11 of library BA0AB from strain CLB 210 of  
DEFINITION Kluyveromyces lactis, sequence tagged site.  
ACCESSION AL428614  
VERSION AL428614.1 GI:12211808  
KEYWORDS STS.  
SOURCE Kluyveromyces lactis.  
ORGANISM Kluyveromyces lactis.  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
REFERENCE  
AUTHORS 1 (bases 1 to 855)  
Bolotin-Fukuhara, M., Toffano-Nioche, C., Artiguenave, F.,  
Duchateau-Nguyen, G., Lemaire, M., Marmisse, R., Montchoer, R.,  
Robert, C., Termier, M., Wincker, P., and Wesolowski-Louvel, M.  
Genomic Exploration of the Hemiascomycetous Yeasts: 11.  
Kluyveromyces lactis  
TITLE  
JOURNAL FEBS Lett. 487 (1), 66-70 (2000)  
PUBMED 11152886  
REFERENCE 2 (bases 1 to 855)  
Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,  
Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,

de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,  
Malpertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Potier, S.,  
Sautin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,  
Wincker, P., and Weissbach, J.  
Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of  
yeast species for molecular evolution studies(1)  
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)  
PUBMED 11152876  
REFERENCE 3 (bases 1 to 855)

AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Genoscope.  
Direct Submission  
Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,  
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
This STS is part of a random genomic sequencing program of fifteen  
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces  
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,  
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces  
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia  
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,  
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to  
5 kb were prepared and both extremities were sequenced. See  
keywords for description of this sequence and for the sequence of  
the other extremity of this insert.

FEATURES  
source  
1. .855  
/organism="Kluyveromyces lactis"  
/strain="CLB 210"  
/variety="lactis"  
/db\_xref="taxon:28985"  
/clone="BA0AB026G11"  
/clone\_lib="BA0AB"

misc\_feature  
<2. .>853  
/note="similar to Saccharomyces cerevisiae ORF YMR186w [HSC82 ; heat shock protein ] similar to Saccharomyces cerevisiae ORF YPL240c [ HSP82 ; heat shock protein ]"  
/evidence="not-experimental"

BASE COUNT 273 a 181 c 180 g 211 t 10 others  
ORIGIN

Query Match 1.8%; Score 19; DB 53; Length 855;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 681 ggaagactctgaacaattc 699  
|||||  
DB 91 GGAAGACTCTGACAAATTC 109

RESULT 43  
OCMP 870 bp mRNA INV 13-JUN-1997  
LOCUS OSTERAGIA circumcincta mRNA for mitochondrial processing  
DEFINITION  
ACCESSION Y13551  
VERSION Y13551.1 GI:2196670  
KEYWORDS mitochondrial processing peptidase; MPP gene.  
SOURCE Teladorsagia circumcincta.  
ORGANISM Teladorsagia circumcincta.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;  
Trichostrongyloidea; Haemonchidae; Osteragiinae; Teladorsagia.  
REFERENCE  
AUTHORS 1 (bases 1 to 870)  
Walker, J. and Tait, A.  
Osteragia circumcincta (Nematoda): Isolation of a partial cDNA  
encoding an unusual member of the mitochondrial processing  
peptidase subfamily of M16 metallopeptidases  
TITLE  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 870)  
Walker, J.  
Direct Submission  
JOURNAL Submitted (02-JUN-1997) J. Walker, University of Glasgow, Infection



REMARK	REFERENCE	TITLE	JOURNAL	COMMENT
and Immunity IRLS, Joseph Black Building, University of Glasgow, Glasgow G12 8QQ, UK	3 (bases 1 to 870)	Walker, J.	Direct Submission	Submitted (13-JUN-1997) J. Walker, University of Glasgow, Infection and Immunity IRLS, Joseph Black Building, University of Glasgow, Glasgow G12 8QQ, UK
On Jun 14, 1997 this sequence version replaced gi:2182026.	Location/Qualifiers			
source	1..870	/organism="Teladorsagia circumcincta"		
		/specific_host="sheep"		
		/strain="Glasgow University Vet. School Strain"		
		/db_xref="taxon:45464"		
		/dev_stage="adult"		
		/lab_host="sheep"		
	25..870	/gene="Mpp"		
	25..>870	/gene="Mpp"		
		/codon_start=1		
		/product="mitochondrial processing peptidase"		
		/protein_id="CAA73887.1"		
		/db_xref="GI:2182027"		
		/db_xref="SPTREMBL:O02378"		
		/translation="MALREYVIGSLAPVLTNPQVNVALSTKDIYLANPCEVHTLKKKLGFRVAADNKOATFVGVWLETSSRTENEGNNGYAHLERLMHKGTGRASAKLSLEPFLAARQSYTRDRTAFVFOSSSEDERVVDILADLVNLSKLSVSEADNRVLLRLEEEAGGQVGMIDMLHLAAVQCTSMKSPGISTSLKISQHLKEMQEDNRTRRVLVAIVGGCGSGSKLOGIAEYFGDLSNEYPRKPEGGIRFGYERYRNDYIPIHYVA"		
		ANALRA"		
BASE COUNT	230 a	191 c	245 g	204 t
ORIGIN				
Query Match	1..88;	Score 19;	DB 96;	Length 870;
Best Local Similarity	100.0%;	Pred. No. 31;		
Matches	19;	Conservative	0;	Mismatches 0;
				Gaps 0;
07	131	acctgcctgcattgcac	149	
Db	86	ACGTGCTGCATTGTCAT	104	
RESULT 44				
AF201698	898 bp	mRNA	02-DEC-2000	
LOCUS	Mus musculus	homeobox protein GPBOX (gpbox)	mRNA, complete cds.	
DEFINITION	AF201698			
ACCESSION	AF201698.1	GI:11526759		
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
		house mouse.		
		Mus musculus		
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 898)			
AUTHORS	Takasaki, N., McIsaac, R. and Dean, J.			
TITLE	Gpbox, a novel homeobox gene preferentially expressed in female germ cells at the onset of sexual dimorphism in mice			
	Unpublished			
JOURNAL	2 (bases 1 to 998)			
REFERENCE	Takasaki, N., McIsaac, R. and Dean, J.			
AUTHORS	Submitted (04-NOV-1999)			
TITLE	Direct Submission			
JOURNAL	Submitted (04-NOV-1999)			
	Biology, NIDDK, National Institutes of Health, Building 6, Room B1-26 9000 Rockville Pike, Bethesda, MD 20892, USA			
FEATURES				
source	1..898			
	/organism="Mus musculus"			
	/db_xref="taxon:10090"			
	1..898			

FEATURES	LOCATION/Qualifiers
CDS	<p>/gene="Gpbpx"</p> <p>92...775</p> <p>/gene="Gpbpx"</p> <p>/codon_start=1</p> <p>/product="homeobox protein Gpbpx"</p> <p>/protein_id="AAC36768.1"</p> <p>/db_xref="GI:11526760"</p> <p>/translation="METPDGDSKQSIQKPPSPGAEDDKERQHGCAANYISGAGEGIDKKELIWMGSLAAGDLOGEAGVAGEQAQAEAPLSPADATGEEGKKBEMGRHAGDCAASPEDDNIQEGGENTIDQPPQDEAAIPEDMRNPQAVYLAHQTRTRPTHSQLRLEETPEENRPFSLRVRRDLARMGVDESDYEWFKMRRLFRHRLAMFCELPTPEENSP"</p>
BASE COUNT	259 a 203 c 273 g 163 t
ORIGIN	
Query Match	1.8%; Score 19; DB 94; Length 896;
Best Local Similarity	100.0%; Pred. No. 31;
Matches	19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy	764 atggtgggaatgcagttggt 782
Db	171 ATGTGGTGAATGCAGTGT 189
RESULT 45	
CNS07120/c	CNS07120 1027 bp DNA STRS 11-JAN-2001
LOCUS	clone BA0AB06601 of library BA0AB from strain CLIB 210 of
DEFINITION	Kluyveromyces lactis, sequence tagged site.
ACCESSION	AL425578
VERSION	AL425578.1 GI:12208772
KEYWORDS	STS.
SOURCE	Kluyveromyces lactis.
ORGANISM	Kluyveromyces lactis
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
AUTHORS	1 (bases 1 to 1027)
TITLE	Bolotin-Fukuhara,M., Toffano-Nioche,C., Artiguenave,F., Duchateau-Neuven,G., Lemaire,M., Marnessee,R., Montroche,R., Robert,C., Termler,M., Winkler,P. and Wesolowski-Louvel,M.
JOURNAL	Genomic Exploration of the Hemiascomycetous Yeasts: 11.
PUBMED	Kluyveromyces lactis
REFERENCE	1152886
AUTHORS	2 (bases 1 to 1027)
TITLE	Soucier,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brotlier,P., Casaregola,S., deMontigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,C., Potier,S., Sautin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Winkler,P. and Weissbach,J.
JOURNAL	Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
PUBMED	yeast species for molecular evolution studies(1)
REFERENCE	1152876
AUTHORS	3 (bases 1 to 1027)
TITLE	Genoscope.
JOURNAL	Direct Submission
PUBMED	Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
REFERENCE	2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
AUTHORS	seqref@genoscope.cns.fr Web : www.genoscope.cns.fr)
TITLE	This STS is part of a random genomic sequencing program of thirteen
JOURNAL	yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
PUBMED	exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
REFERENCE	Saccharomyces kluyveri, Kluyveromyces marxianus var. marxianus, Pichia
AUTHORS	lactis var. lactis, Kluyveromyces hanseni var. hanseni, Pichia
TITLE	angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
JOURNAL	Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
PUBMED	5 kb were prepared and both extremities were sequenced. See
REFERENCE	keywords for description of this sequence and for the sequence of
AUTHORS	the other extremity of this insert.
TITLE	location/Qualifiers

```

source
1..1027
/organism="Kluyveromyces lactis"
/strain="CLIB 210"
/variety="lactis"
/db_xref="taxon:28995"
/clone="BA0AB06G01"
/clone_1ib="BA0AB"
complement(<3..>1018)
/note="similar to Saccharomyces cerevisiae ORF YMR186w [
HSC82 ; heat shock protein ]
similar to Saccharomyces cerevisiae ORF YPL240c [ HSP82 ;
heat shock protein ]"
/evidence="not_experimental"

BASE COUNT      248 a      238 c      189 g      350 t      2 others
ORIGIN

Query Match      1.8%; Score 19; DB 53; Length 1027;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      681 ggaagacttaaacattc 699
|||||
Db      60 ggaagacttaaacattc 42

RESULT 46
AB025275      1439 bp      mRNA      INV      06-AUG-1999
LOCUS      Stichopus japonicus mRNA for arginine kinase, complete cds.
DEFINITION      AB025275
VERSION      AB025275.1 GI:4586461
KEYWORDS      arginine kinase.
SOURCE      Stichopus japonicus cDNA to mRNA.
ORGANISM      Stichopus japonicus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Holothuroidea; Aspidochirotea; Aspidochirotida; Stichopodidae;
Stichopus.
1 (sites)
Suzuki,T., Kamidochi,M., Inoue,N., Kawamichi,H., Yazawa,Y.,
Furukohji,T. and Ellington,R.W.
Arginine Kinase Evolved Twice: Evidence That Echinoderm Arginine
Kinase Originated From Creatine Kinase
Unpublished (1999)
2 (bases 1 to 1439)
Suzuki,T.
Direct Submission
Submitted (24-MAR-1999) to the DDBJ/EMBL/GenBank databases.
Memoriko Suzuki, Kochi University, Faculty of Science, Akebonocho
2-5-1, Kochi, Kochi 780-8520, Japan
(E-mail:suzuki@sc.kochi-u.ac.jp, Tel:81-888-44-0111(ex.1264),
Fax:81-888-44-8356)
Location/Qualifiers
1..1439
/organism="Stichopus japonicus"
/db_xref="taxon:7692"
27..1139
/codon_start=1
/product="arginine kinase"
/protein_id="BA06385.1"
/db_xref="GI:4586462"
/translation="MANLNOKKYPKADDPNFEGHSLSKYLITADMYAKIDAVTPS
GYTLIDRAIONGVNDPFLGLAGDETVYRADLPDPIVEIHNGFKTDNHTDLD
ASKTLDVLDPAVVISRVPTGRIRRMALSHVCSRRRAIEVWSEALNSLAADLK
GKYSYLMKMDERTQOOLIDHFLDFRVSHTSGMGRDPPGRGIWINDKNFLVW
INEDHTRITISQMGNNKVEFEFTRGLEVEKIKDKGKPEKMNNDLGFVLTCPS
NLGTVCSYHAKLPHMAKDKREFEICTKMLQKRGTSGETPSVGVYDISNDRIC
SSVEQVNCYIKGVVLLEMEKKLEGGESIDLVPR"

BASE COUNT      417 a      303 c      335 g      384 t

Query Match      1.8%; Score 19; DB 4; Length 1439;

```

```

Best Local Similarity 100.0%; Pred. No. 31;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      703 gatgcttaagaccttg 721
|||||
Db      330 gatgcttaagaccttg 348

RESULT 47
SCYLRO12C/c      1660 bp      DNA      PLN      11-AUG-1997
LOCUS      S.cerevisiae chromosome XII reading frame ORF YLR012C.
DEFINITION      273184 Y13138
ACCESSION      273184.1 GI:1360307
VERSION      273184.1 GI:1360307
KEYWORDS
SOURCE      Saccharomyces cerevisiae.
ORGANISM      Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
1 (bases 1 to 1660)
Vandenbol,M., Portetelle,D. and Hilger,F.
Unpublished
2 (bases 1 to 1660)
MIPS.
Direct Submission
Submitted (22-MAY-1996) Data collected by MIPS on behalf of the
European yeast chromosome XII sequencing project...MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
Martinsried, FRG; E-mail: Mewes@mips.emblnet.org
Location/Qualifiers
1..1660
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/chromosome="XII"
/complement(304..672)
/note="ORF YLR012c"
/codon_start=1
/protein_id="CAA97534.1"
/db_xref="GI:1360308"
/translation="MTKASQKLFIEIKRSLEHIOKTYVYKEITTYOYLQLOPQ
OEKTLALCOQDEQETRIAFDRGGVPGIARKFAQEEVAMFDVVTWSYMAVTPSY
RRRNRLKIDMLKMSNAEY"

BASE COUNT      536 a      299 c      264 g      561 t

Query Match      1.8%; Score 19; DB 15; Length 1660;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      889 tataagataatttgat 907
|||||
Db      965 tataagataatttgat 947

RESULT 48
SCYLRO13W/c      1917 bp      DNA      PLN      11-AUG-1997
LOCUS      S.cerevisiae chromosome XII reading frame ORF YLR013W.
DEFINITION      273185 Y13138
ACCESSION      273185.1 GI:1360309
VERSION      273185.1 GI:1360309
KEYWORDS
SOURCE      Saccharomyces cerevisiae.
ORGANISM      Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
1 (bases 1 to 1917)
Vandenbol,M., Portetelle,D. and Hilger,F.
Unpublished
2 (bases 1 to 1917)
MIPS.
Direct Submission

```

## JOURNAL

Submitted (22-MAY-1996) Data collected by MIPS on behalf of the European yeast chromosome XII sequencing project. MIPS at the Max Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152 Martinsried, FRG; E-mail: Mewes@mips.embl.org

## FEATURES

Location/Qualifiers

## CDS

1..1917  
/organism="Saccharomyces cerevisiae"  
/db\_xref="taxon:4932"  
/chromosome="XII"  
989..1414  
/note="ORF YLR013w"  
/codon\_start=1  
/protein\_id="CAA97535.1"  
/db\_xref="GI:1360310"  
/db\_xref="SPTREMBL:O07928"  
/translation="MNKIKCHPEYKRISVESILNPVEETIDCKPHSOTKINTAKPI  
SASLVYNNNAVVOHNOVKRGYTRCPCQCAVITSPOMREGPDGVTILCANAGLIFY  
RRIFLVGGKLAKRYNEIKGVSKRVRPKSLIGVTRIR"  
BASE COUNT 639 a 335 c 334 g 609 t  
ORIGIN

## Query Match

Best Local Similarity 100.0%; Pred. No. 31;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 889 tataagataatttgaat 907  
|||||  
Db 293 TATAAGTATAATTGCAAT 275

## RESULT 49

AC014479/c 25818 bp DNA HTG 16-NOV-1999

LOCUS Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, In ordered  
pieces.

ACCESSION AC014479  
VERSION AC014479.1 GI:6436856  
KEYWORDS HTGS\_PHASE2.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 25818)  
Adams, M. and Venter, J. C.  
Direct Submission

Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,  
Rockville, MD, USA

## COMMENT

This sequence was identified as CDM:10214121 by the submitter.  
For further information on this sequence e-mail to fly@celera.com.

\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

## FEATURES

source

1..25818  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
BASE COUNT 7271 a 5623 c 5714 g 7210 t  
ORIGIN

Query Match 1.8%; Score 19; DB 63; Length 25818;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 cctcaattctctgtagag 525  
|||||  
Db 14647 CCTCAATTCCTGAGTAGG 14629

## RESULT 50

## HS1168A5

LOCUS HS1168A5 75522 bp DNA PRI 23-NOV-1999  
DEFINITION Human DNA sequence from clone 1168A5 on chromosome Xq23 Contains  
GSS, complete sequence.

ACCESSION AL031183  
VERSION AL031183.4 GI:4468272  
KEYWORDS HTG.

## SOURCE

ORGANISM

## REFERENCE

1 (bases 1 to 75522)  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

## AUTHORS

JOURNAL

## COMMENT

On Mar 22, 1999 this sequence version replaced gi:4467177.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
IMPORTANT: This sequence is not the entire insert of clone 1168A5.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we arrange for a small overlap between  
neighbouring submissions.  
The true left end of clone 1170D6 is at 75423 in this sequence.  
This sequence has been finished according to sequence map criteria  
as follows. An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'  
feature key.  
This sequence was generated from part of bacterial clone contigs of  
human chromosome X, constructed by the Sanger Centre Chromosome X  
Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/chrX>  
1168A5 is from the library RPI5 constructed at the Roswell Park  
Cancer Institute by the group of Pieter de Jong. For further  
details see <http://bacpac.med.buffalo.edu/VECTOR:pcyPAC2>.

## FEATURES

source

1..75522  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/map="q23"  
/clone.lib="RPI-5"  
1..660  
/note="LIP repeat: matches 2..667 of consensus"  
666..2100  
/note="LIP repeat: matches 1670..3097 of consensus"  
2079..3823  
/note="LIP repeat: matches 1670..3097 of consensus"  
4454..4849  
/note="LIP repeat: matches 1670..3097 of consensus"  
5948..6049  
/note="LIP repeat: matches 1670..3097 of consensus"  
6394..6721  
/note="LIP repeat: matches 1670..3097 of consensus"  
8984..9280  
/note="LIP repeat: matches 1670..3097 of consensus"  
9748..9790  
/note="LIP repeat: matches 1670..3097 of consensus"  
9796..10054  
/note="LIP repeat: matches 1670..3097 of consensus"  
10077..10630  
/note="LIP repeat: matches 1670..3097 of consensus"  
11033..11469  
/note="LIP repeat: matches 1670..3097 of consensus"  
12177..12509  
/note="LIP repeat: matches 1670..3097 of consensus"

```

repeat_region /note="AluIb repeat: matches 1. .302 of consensus"
16334. .16810
/note="L1ME3 repeat: matches 5699. .6155 of consensus"
repeat_region /note="L1ME3 repeat: matches 5699. .6155 of consensus"
16812. .16886
/note="MER69 repeat: matches 1. .75 of consensus"
repeat_region /note="L1MA2 repeat: matches 6052. .6302 of consensus"
16968. .17214
/note="L1PA7 repeat: matches 3997. .6134 of consensus"
repeat_region /note="L1PA7 repeat: matches 3997. .6134 of consensus"
17227. .19327
19325. .19852
/note="L1P repeat: matches 2650. .3176 of consensus"
repeat_region /note="L1P repeat: matches 2650. .3176 of consensus"
19864. .23603
/note="L1M1 repeat: matches -1186. .2650 of consensus"
misc_feature /note="L1M1 repeat: matches -1186. .2650 of consensus"
21770. .>22001
/note="match: STS L42664"
repeat_region /note="match: STS L42664"
23576. .24214
/note="L1M1 repeat: matches -1390. .-743 of consensus"
repeat_region /note="L1M1 repeat: matches -1390. .-743 of consensus"
24447. .25183
/note="L1MC/D repeat: matches 4942. .5703 of consensus"
repeat_region /note="L1MC/D repeat: matches 4942. .5703 of consensus"
25286. .25566
/note="L1M4 repeat: matches 4039. .4331 of consensus"
repeat_region /note="L1M4 repeat: matches 4039. .4331 of consensus"
25574. .25825
/note="L12 repeat: matches 2457. .2750 of consensus"
repeat_region /note="L12 repeat: matches 2457. .2750 of consensus"
25879. .26049
/note="MIR repeat: matches 50. .190 of consensus"
repeat_region /note="MIR repeat: matches 50. .190 of consensus"
26203. .26300
/note="AluIy repeat: matches 214. .309 of consensus"
repeat_region /note="AluIy repeat: matches 214. .309 of consensus"
26760. .27527
/note="L1M3A repeat: matches 5527. .6098 of consensus"
repeat_region /note="L1M3A repeat: matches 5527. .6098 of consensus"
28565. .29134
/note="L1M3A repeat: matches 5175. .5399 of consensus"
repeat_region /note="L1M3A repeat: matches 5175. .5399 of consensus"
29217. .29424
/note="L1M3A repeat: matches 4826. .6161 of consensus"
repeat_region /note="L1M3A repeat: matches 4826. .6161 of consensus"
29426. .30746
/note="L1PA8 repeat: matches 4814. .5019 of consensus"
repeat_region /note="L1PA8 repeat: matches 4814. .5019 of consensus"
30877. .31089
/note="L1M3A repeat: matches 3. .311 of consensus"
repeat_region /note="L1M3A repeat: matches 3. .311 of consensus"
31090. .31387
/note="AluSp repeat: matches 3. .311 of consensus"
repeat_region /note="AluSp repeat: matches 3. .311 of consensus"
31388. .31500
/note="L1M3A repeat: matches 4700. .4814 of consensus"
repeat_region /note="L1M3A repeat: matches 4700. .4814 of consensus"
31978. .32072
/note="MER5A repeat: matches 94. .188 of consensus"
repeat_region /note="MER5A repeat: matches 94. .188 of consensus"
32082. .32255
/note="MER20 repeat: matches 1. .218 of consensus"
repeat_region /note="MER20 repeat: matches 1. .218 of consensus"
32768. .33125
/note="THE1B repeat: matches 1. .364 of consensus"
repeat_region /note="THE1B repeat: matches 1. .364 of consensus"
<33104. .>33606
/note="match: GSS A0169991"
repeat_region /note="match: GSS A0169991"
33221. .33290
/note="MIR repeat: matches 71. .142 of consensus"
repeat_region /note="MIR repeat: matches 71. .142 of consensus"
34200. .34702
/note="L1MB8 repeat: matches 5697. .6164 of consensus"
repeat_region /note="L1MB8 repeat: matches 5697. .6164 of consensus"
34703. .35250
/note="L1PA8 repeat: matches 5612. .6159 of consensus"
repeat_region /note="L1PA8 repeat: matches 5612. .6159 of consensus"
35251. .35729
/note="L1MB8 repeat: matches 5208. .5697 of consensus"
repeat_region /note="L1MB8 repeat: matches 5208. .5697 of consensus"
37283. .38025
/note="L1PB3 repeat: matches 3205. .3951 of consensus"
repeat_region /note="L1PB3 repeat: matches 3205. .3951 of consensus"
38022. .39658
/note="L1PB3 repeat: matches 3951. .5603 of consensus"
repeat_region /note="L1PB3 repeat: matches 3951. .5603 of consensus"
39651. .40092
/note="L1PB3 repeat: matches 5704. .6150 of consensus"
repeat_region /note="L1PB3 repeat: matches 5704. .6150 of consensus"
40100. .40137
/note="19 copies 2 mer aa 84% conserved"
repeat_region /note="19 copies 2 mer aa 84% conserved"
41228. .41444
/note="MIR repeat: matches 32. .257 of consensus"
repeat_region /note="MIR repeat: matches 32. .257 of consensus"
41519. .41677
/note="MER5A repeat: matches 29. .189 of consensus"
repeat_region /note="MER5A repeat: matches 29. .189 of consensus"
41593. .>41888
/note="match: GSS A0056229 clone 2339E7"
misc_feature /note="match: GSS A0056229 clone 2339E7"
41886. .42441
/note="match: GSS B68354 clone 2024D6"

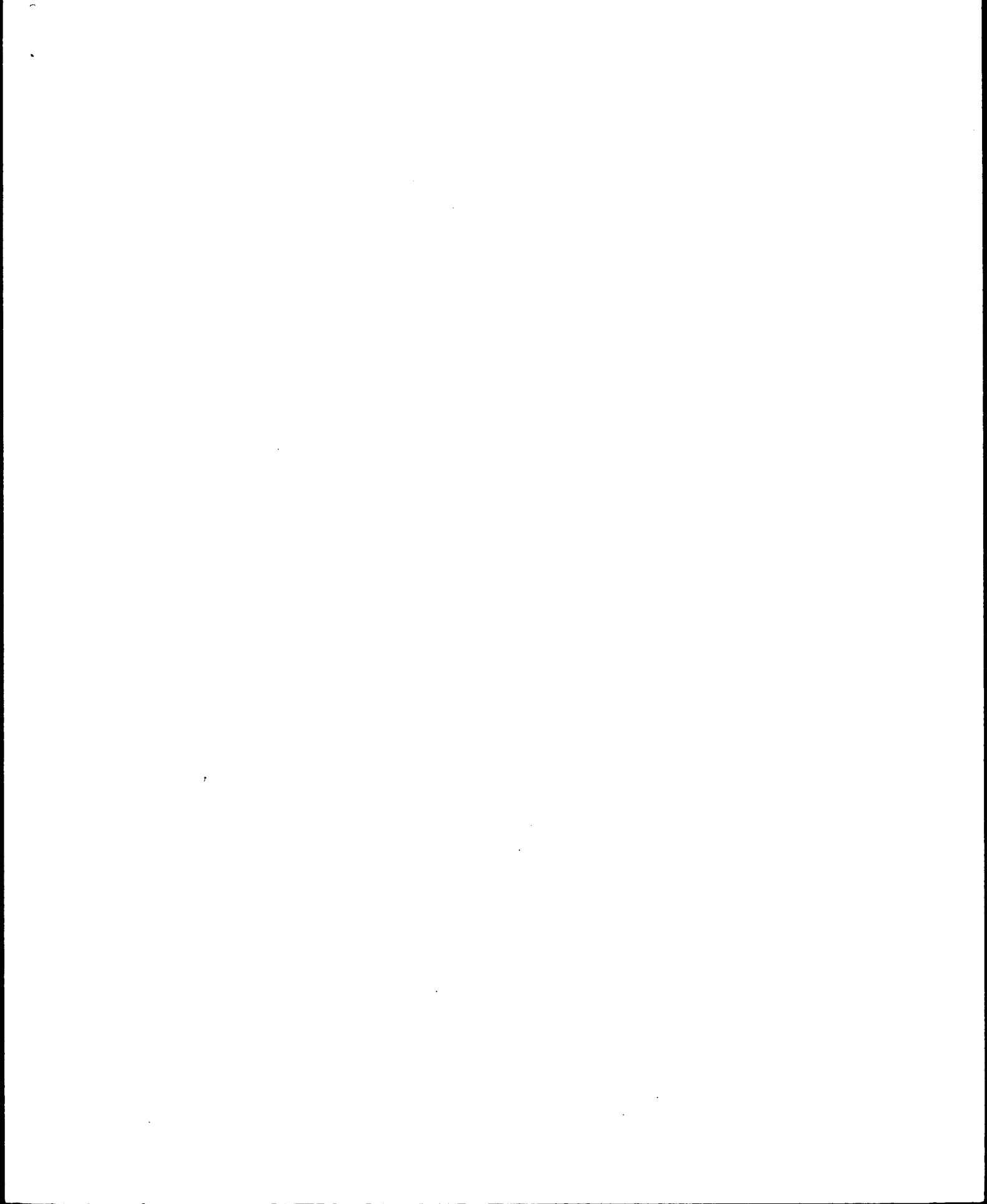
```

Search completed: August 19, 2001, 01:09:16  
Job time: 17536 sec

Query Match 1.8%; Score 19; DB 92; Length 75522;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0Y 483 aaactgtcttcagttca 501  
|||||  
DB 66312 AACGCTGTCTCAGTTCA 66330





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 17, 2001, 16:43:43 ; Search time 12.41 Seconds

(without alignments)  
560,711 Million cell updates/sec

Title: US-09-284-320-6

Perfect score: 350

Sequence: 1 MAFFVVLALVAGVIGNEFS.....MDPGYSITRYMTNOKIRMD 350

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 197339 seqs, 20590346 residues

Word size : 5

Total number of hits satisfying chosen parameters: 5746

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCNUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfill1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	74	21.1	93	4 US-08-905-223-402	Sequence 402, App
2	7	2.0	93	1 US-08-464-339A-13	Sequence 13, Appl
3	7	2.0	95	4 US-08-654-482-9	Sequence 9, Appl
4	7	2.0	103	2 US-08-796-414B-3	Sequence 3, Appl
5	7	2.0	137	4 US-09-091-117-4	Sequence 4, Appl
6	7	2.0	348	1 US-08-468-847B-14	Sequence 14, Appl
7	7	2.0	349	1 US-08-167-628-2	Sequence 2, Appl
8	7	2.0	349	1 US-08-386-680-2	Sequence 2, Appl
9	7	2.0	349	1 US-08-459-717-2	Sequence 2, Appl
10	7	2.0	349	1 US-08-712-302-2	Sequence 2, Appl
11	7	2.0	349	2 US-08-880-031-2	Sequence 2, Appl
12	7	2.0	349	2 US-09-054-368-2	Sequence 2, Appl
13	7	2.0	349	4 US-09-097-179-2	Sequence 2, Appl
14	7	2.0	349	4 US-09-054-274-2	Sequence 2, Appl
15	7	2.0	349	4 US-09-080-715-2	Sequence 2, Appl
16	7	2.0	349	4 US-09-056-704-2	Sequence 2, Appl
17	7	2.0	349	5 PCT-US96-08140-2	Sequence 2, Appl
18	7	2.0	377	4 US-08-969-106-6	Sequence 6, Appl
19	7	2.0	377	4 US-09-054-492B-1	Sequence 1, Appl
20	7	2.0	401	1 US-08-368-803-7	Sequence 7, Appl
21	7	2.0	422	1 US-08-663-566A-5	Sequence 5, Appl
22	7	2.0	422	2 US-08-023-610-5	Sequence 5, Appl
23	7	2.0	422	2 US-08-288-065A-5	Sequence 5, Appl
24	7	2.0	422	2 US-08-362-240A-5	Sequence 5, Appl
25	7	2.0	422	2 US-08-362-372A-3	Sequence 5, Appl
26	7	2.0	422	5 PCT-US95-10245-5	Sequence 5, Appl
27	7	2.0	450	2 US-08-611-280-2	Sequence 2, Appl

28	7	2.0	620	4 US-09-000-145-3	Sequence 3, Appl
29	7	2.0	841	4 US-09-413-814-107	Sequence 107, App
30	7	2.0	2618	4 US-09-413-814-28	Sequence 28, Appl
31	6	1.7	9	3 US-08-159-339A-917	Sequence 917, App
32	6	1.7	13	1 US-07-958-903A-18	Sequence 18, Appl
33	6	1.7	13	1 US-08-462-018-18	Sequence 18, Appl
34	6	1.7	13	1 US-08-823-245-18	Sequence 18, Appl
35	6	1.7	13	5 PCT-US92-09443A-18	Sequence 18, Appl
36	6	1.7	15	2 US-08-350-260A-46	Sequence 46, Appl
37	6	1.7	15	2 US-08-350-260A-48	Sequence 48, Appl
38	6	1.7	17	2 US-08-637-759B-65	Sequence 65, Appl
39	6	1.7	17	3 US-08-871-355A-65	Sequence 65, Appl
40	6	1.7	20	2 US-08-484-530-34	Sequence 34, Appl
41	6	1.7	20	2 US-08-827-618A-34	Sequence 34, Appl
42	6	1.7	20	3 US-08-483-952A-34	Sequence 34, Appl
43	6	1.7	24	5 US-08-602-999A-143	Sequence 143, App
44	6	1.7	24	5 PCT-US92-09443A-73	Sequence 73, Appl
45	6	1.7	28	1 US-08-090-036-2	Sequence 2, Appl
46	6	1.7	29	1 US-08-435-252-1	Sequence 1, Appl
47	6	1.7	29	5 PCT-US96-01314-68	Sequence 68, Appl
48	6	1.7	30	1 US-08-066-325-72	Sequence 72, Appl
49	6	1.7	31	1 US-08-090-036-1	Sequence 1, Appl
50	6	1.7	31	1 US-08-190-802A-247	Sequence 247, App
51	6	1.7	31	4 US-08-681-838A-5	Sequence 5, Appl
52	6	1.7	33	6 5470721-2	Patent No. 5470721
53	6	1.7	34	4 US-08-905-223-342	Sequence 342, App
54	6	1.7	36	6 5489517-2	Patent No. 5489517
55	6	1.7	38	1 US-08-215-084A-7	Sequence 7, Appl
56	6	1.7	38	1 US-08-463-212-7	Sequence 7, Appl
57	6	1.7	38	1 US-08-463-211-7	Sequence 7, Appl
58	6	1.7	39	2 US-08-350-260A-52	Sequence 52, Appl
59	6	1.7	40	1 US-08-188-228-20	Sequence 20, Appl
60	6	1.7	40	1 US-08-332-643-20	Sequence 20, Appl
61	6	1.7	40	1 US-08-332-638-20	Sequence 20, Appl
62	6	1.7	50	6 5436136-16	Patent No. 5436136
63	6	1.7	51	1 US-08-208-108-11	Sequence 11, Appl
64	6	1.7	52	4 US-09-346-860-8	Sequence 8, Appl
65	6	1.7	57	2 US-08-480-229C-23	Sequence 23, Appl
66	6	1.7	57	2 US-08-659-235C-23	Sequence 23, Appl
67	6	1.7	59	2 US-09-006-675-4	Sequence 4, Appl
68	6	1.7	59	4 US-09-300-672-8	Sequence 8, Appl
69	6	1.7	65	2 US-07-808-457-9	Sequence 9, Appl
70	6	1.7	65	5 PCT-US92-10178-9	Sequence 9, Appl
71	6	1.7	67	5 PCT-US92-09443A-2	Sequence 2, Appl
72	6	1.7	68	3 US-08-912-272-26	Sequence 26, Appl
73	6	1.7	70	1 US-07-654-611-2	Sequence 2, Appl
74	6	1.7	70	1 US-08-180-572-5	Sequence 5, Appl
75	6	1.7	70	1 US-07-947-035-1	Sequence 1, Appl
76	6	1.7	70	1 US-07-776-272-17	Sequence 17, Appl
77	6	1.7	70	1 US-07-958-903A-17	Sequence 17, Appl
78	6	1.7	70	1 US-08-462-018-17	Sequence 17, Appl
79	6	1.7	70	1 US-08-823-245-17	Sequence 17, Appl
80	6	1.7	70	1 US-08-482-271-1	Sequence 1, Appl
81	6	1.7	70	3 US-09-003-708A-4	Sequence 4, Appl
82	6	1.7	70	3 US-09-080-120A-1	Sequence 1, Appl
83	6	1.7	70	3 US-08-432-517-1	Sequence 1, Appl
84	6	1.7	70	3 US-09-383-212-4	Sequence 4, Appl
85	6	1.7	70	5 PCT-US92-09443A-1	Sequence 1, Appl
86	6	1.7	70	5 PCT-US93-11458-1	Sequence 1, Appl
87	6	1.7	70	5 PCT-US95-08925-1	Sequence 1, Appl
88	6	1.7	76	1 US-08-338-558-1	Patent No. 5470828
89	6	1.7	76	1 US-08-460-890A-47	Sequence 1, Appl
90	6	1.7	78	2 US-08-167-641C-47	Sequence 47, Appl
91	6	1.7	78	4 US-08-460-971A-47	Sequence 47, Appl
92	6	1.7	78	4 US-08-462-040-47	Sequence 47, Appl
93	6	1.7	81	4 US-08-905-223-348	Sequence 348, App
94	6	1.7	82	4 US-08-905-223-449	Sequence 449, App
95	6	1.7	83	1 US-07-947-035-18	Sequence 18, Appl
96	6	1.7	83	1 US-08-321-585A-12	Sequence 12, Appl
97	6	1.7	86	2 US-08-744-670-1	Sequence 1, Appl
98	6	1.7	86	2 US-09-149-933-1	Sequence 1, Appl
99	6	1.7	91	2 US-07-808-457-21	Sequence 21, Appl
100	6	1.7	91	2	















```

977 5 1.4 35 3 US-08-484-223B-53 Sequence 53, Appl
978 5 1.4 35 3 US-08-484-223B-147 Sequence 147, App
979 5 1.4 35 3 US-08-919-597-49 Sequence 49, Appl
980 5 1.4 35 3 US-08-919-597-50 Sequence 50, Appl
981 5 1.4 35 3 US-08-919-597-51 Sequence 51, Appl
982 5 1.4 35 3 US-08-919-597-52 Sequence 52, Appl
983 5 1.4 35 3 US-08-919-597-53 Sequence 53, Appl
984 5 1.4 35 3 US-08-919-597-147 Sequence 147, App
985 5 1.4 35 3 US-08-475-668A-49 Sequence 49, Appl
986 5 1.4 35 3 US-08-475-668A-50 Sequence 50, Appl
987 5 1.4 35 3 US-08-475-668A-51 Sequence 51, Appl
988 5 1.4 35 3 US-08-475-668A-52 Sequence 52, Appl
989 5 1.4 35 3 US-08-475-668A-53 Sequence 53, Appl
990 5 1.4 35 3 US-08-475-668A-147 Sequence 147, App
991 5 1.4 35 3 US-08-485-551A-49 Sequence 49, Appl
992 5 1.4 35 3 US-08-485-551A-50 Sequence 50, Appl
993 5 1.4 35 3 US-08-485-551A-51 Sequence 51, Appl
994 5 1.4 35 3 US-08-485-551A-52 Sequence 52, Appl
995 5 1.4 35 3 US-08-485-551A-53 Sequence 53, Appl
996 5 1.4 35 3 US-08-485-551A-147 Sequence 147, App
997 5 1.4 35 3 US-08-471-913A-49 Sequence 49, Appl
998 5 1.4 35 3 US-08-471-913A-50 Sequence 50, Appl
999 5 1.4 35 3 US-08-471-913A-51 Sequence 51, Appl
1000 5 1.4 35 3 US-08-471-913A-52 Sequence 52, Appl

```

## ALIGNMENTS

```

RESULT 1
US-08-905-223-402
; Sequence 402, Application US/08905223
; Patent No. 6222029
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste D.
; APPLICANT: Duclert, Aymeric
; APPLICANT: Lacroix, Bruno
; TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS
; NUMBER OF SEQUENCES: 503
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knodde, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/905,223
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 402:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; TISSUE TYPE: Brain
; FEATURE:
; NAME/KEY: sig_peptide

```

```

; LOCATION: -16..-1
; IDENTIFICATION METHOD: Von Heijne matrix
; OTHER INFORMATION: score 12.5
; OTHER INFORMATION: seq FVLLALVAVGLG/NE
US-08-905-223-402

```

```

Query Match 21.1%; Score 74; DB 4; Length 93;
Best Local Similarity 100.0%; Pred. No. 1, Be-61;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 MAVFVLLALVAVGLGNEFSILKSPGVFRNGNMPDGERIPVVALSMGSFVEDLSW 60
Db 1 MAVFVLLALVAVGLGNEFSILKSPGVFRNGNMPDGERIPVVALSMGSFVEDLSW 60
OY 61 PGLAVGNLFHRPRA 74
Db 61 PGLAVGNLFHRPRA 74

```

```

RESULT 2
US-08-464-339A-13
; Sequence 13, Application US/08464339A
; Patent No. 5747280
; GENERAL INFORMATION:
; APPLICANT: HASTINGS, ET AL.
; TITLE OF INVENTION: Human Vascular IBP-Like Growth
; FACTOR
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,339A
; FILING DATE: June 5, 1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/14388
; FILING DATE: 9 DEC 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: MILLINS, J G.
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 325800-332
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-08-464-339A-13

```

```

Query Match 2.0%; Score 7; DB 1; Length 93;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 4 FVLLAL 10
Db 13 FVLLAL 19

```

RESULT 3  
US-08-654-482-9  
; Sequence 9, Application US/08654482  
; Patent No. 6245562  
; GENERAL INFORMATION:  
; APPLICANT: Dalia-Favera, Rocco  
; TITLE OF INVENTION: IDENTIFICATION OF GENES ALTERED IN  
; TITLE OF INVENTION: MULTIPLE MYELOMA  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/654,482  
; FILING DATE: 28-MAY-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 50995  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 278-0400  
; TELEFAX: (212) 391-0525  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 95 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-654-482-9

Query Match 2.0%; Score 7; DB 4; Length 95;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 183 FISELOV 189  
|||||  
DB 38 FISELOV 44

RESULT 4  
US-08-796-414B-3  
; Sequence 3, Application US/08796414B  
; Patent No. 5876987  
; GENERAL INFORMATION:  
; APPLICANT: Wendy C. Champness, Paul Brian  
; APPLICANT: and Todd B. Anderson  
; TITLE OF INVENTION: METHOD, DNA AND BACTERIA  
; TITLE OF INVENTION: FOR HYPERPRODUCTION OF AN  
; TITLE OF INVENTION: ANTIBIOTIC DUE TO  
; TITLE OF INVENTION: DISRUPTION OF AN ABSA  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ian C. McLeod  
; STREET: 2190 Commons Parkway  
; CITY: Okemos  
; STATE: Michigan  
; COUNTRY: USA  
; ZIP: 48864

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch,  
; COMPUTER: Acer  
; OPERATING SYSTEM: MS-DOS 5.00  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/796,414B  
; FILING DATE: February 6, 1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ian C. McLeod  
; REGISTRATION NUMBER: 20,931  
; REFERENCE/DOCKET NUMBER: MSU 4.1-297  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (517) 347-4100  
; TELEFAX: (517) 347-4103  
; TELEX: NO. 5876987e  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 103  
; TYPE: Amino Acid  
; STRANDEDNESS: Single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Amino Acid  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE:  
; ORIGINAL SOURCE:  
; ORGANISM: Streptomyces coelicolor  
; STRAIN:  
; INDIVIDUAL ISOLATE:  
; DEVELOPMENTAL STAGE:  
; HAPLOTYPE:  
; TISSUE TYPE:  
; CELL TYPE: unicellular microorganisms  
; CELL LINE:  
; ORGANELLE:  
; IMMEDIATE SOURCE:  
; POSITION IN GENOME:  
; FEATURE:  
; NAME/KEY: D7  
; LOCATION:  
; IDENTIFICATION METHOD: deduced  
; OTHER INFORMATION:  
; PUBLICATION INFORMATION:  
; AUTHORS:  
; TITLE:  
; JOURNAL:  
; VOLUME:  
; ISSUE:  
; PAGES:  
; DATE:  
; RELEVANT RESIDUES IN SEQ ID NO:  
US-08-796-414B-3

Query Match 2.0%; Score 7; DB 2; Length 103;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 104 FSLDSVA 110  
|||||  
DB 82 FSLDSVA 88

RESULT 5  
US-09-091-117-4  
; Sequence 4, Application US/09091117  
; Patent No. 6171589  
; GENERAL INFORMATION:

APPLICANT: The University of Melbourne  
TITLE OF INVENTION: Mycoplasma Recombinant Polypeptides and  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GREENE, WINNER and SULLIVAN P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: United States of America  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/091,117  
FILING DATE: 12 JUNE 1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/AU96/00803  
FILING DATE: 13-DEC-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: AU PN7127  
FILING DATE: 13-DEC-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: WINNER, Ellen P.  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: +1 303 499 8080  
TELEFAX: +1 303 499 8089  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 137 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Mycoplasma genitalium  
US-09-091-117-4

Query Match 2.0%; Score 7; DB 4; Length 137;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 LFLSEL 187  
DB 14 LFLSEL 20

RESULT 6  
US-08-468-847B-14  
Sequence 14, Application US/0846847B  
Patent No. 5780263  
GENERAL INFORMATION:  
APPLICANT: Hastings, Gregg A. and Adams, Mark D.  
TITLE OF INVENTION: Human Con-Like Growth Factor  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
ADDRESSER: CECCHI, STEWART & OLSTEIN  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/468,847B

FILING DATE: 6 June 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: MILLINS, J.G.  
REGISTRATION NUMBER: 33,073  
REFERENCE/DOCKET NUMBER: 325800-442  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 348 AMINO ACIDS  
TYPE: AMINO ACID  
STRANDEDNESS:  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
US-08-468-847B-14

Query Match 2.0%; Score 7; DB 1; Length 348;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FVLLAL 10  
DB 13 FVLLAL 19

RESULT 7  
US-08-167-628-2  
Sequence 2, Application US/08167628  
Patent No. 5408040  
GENERAL INFORMATION:  
APPLICANT: Grotendorst, Gary R.  
TITLE OF INVENTION: CONNECTIVE TISSUE GROWTH FACTOR  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Spensley Horn Jubb & Lubitz  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/167,628  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/752,427  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Weatherell, Jt. Ph.D., John W.  
REGISTRATION NUMBER: 31,678  
REFERENCE/DOCKET NUMBER: PD-1294  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-455-5100  
TELEFAX: 619-455-5100  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 349 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-167-628-2



Query Match 2.0%; Score 7; DB 1; Length 349;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FVLLAL 10  
DB 13 FVLLAL 19

RESULT 8  
US-08-386-680-2  
Sequence 2, Application US/08386680  
Patent No. 5585270  
GENERAL INFORMATION:  
APPLICANT: Grotendorst, Gary R.  
APPLICANT: Bradham Jr., Douglas M.,  
TITLE OF INVENTION: CONNECTIVE TISSUE GROWTH FACTOR  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Spensley Horn Judas & Lubitz  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/386,680  
FILING DATE: 10-FEB-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/167,628  
FILING DATE:  
APPLICATION NUMBER: US/07/752,427  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Wetherell, Jr. Ph.D., John W.  
REGISTRATION NUMBER: 31,678  
REFERENCE/DOCKET NUMBER: PD-1294  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-455-5100  
TELEFAX: 619-455-5110  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 349 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-386-680-2

Query Match 2.0%; Score 7; DB 1; Length 349;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FVLLAL 10  
DB 13 FVLLAL 19

RESULT 9  
US-08-459-717-2  
Sequence 2, Application US/08459717  
Patent No. 5770209  
GENERAL INFORMATION:  
APPLICANT: Grotendorst, Gary R.  
APPLICANT: Bradham Jr., Douglas M.,  
TITLE OF INVENTION: CONNECTIVE TISSUE GROWTH FACTOR  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Spensley Horn Judas & Lubitz  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/712,302  
FILING DATE: 11-SEP-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/386,680

TITLE OF INVENTION: CONNECTIVE TISSUE GROWTH FACTOR  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Spensley Horn Judas & Lubitz  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/459,717  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/752,427  
FILING DATE: 30-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Wetherell, Jr. Ph.D., John W.  
REGISTRATION NUMBER: 31,678  
REFERENCE/DOCKET NUMBER: PD-1294  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-455-5100  
TELEFAX: 619-455-5110  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 349 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-459-717-2

Query Match 2.0%; Score 7; DB 1; Length 349;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FVLLAL 10  
DB 13 FVLLAL 19

RESULT 10  
US-08-712-302-2  
Sequence 2, Application US/08712302  
Patent No. 5783187  
GENERAL INFORMATION:  
APPLICANT: Grotendorst, Gary R.  
APPLICANT: Bradham Jr., Douglas M.,  
TITLE OF INVENTION: CONNECTIVE TISSUE GROWTH FACTOR  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Spensley Horn Judas & Lubitz  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/712,302  
FILING DATE: 11-SEP-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/386,680

FILING DATE: 10-FEB-1995  
APPLICATION NUMBER: US/08/167,628  
FILING DATE:  
APPLICATION NUMBER: US/07/752,427  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Wetherell, Jr. Ph.D., John W.  
REGISTRATION NUMBER: 31,678  
REFERENCE/DOCKET NUMBER: PD-1294  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-455-5100  
TELEFAX: 619-455-5110  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 349 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-712-302-2

Query Match 2.0%; Score 7; DB 1; Length 349;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FVLLAL 10  
|||||||  
DB 13 FVLLAL 19

RESULT 11  
US-08-880-031-2  
Sequence 2, Application US/08880031  
Patent No. 5916756  
GENERAL INFORMATION:  
APPLICANT: Grotendorst, Gary R.  
APPLICANT: Bradham Jr., Douglas M.,  
TITLE OF INVENTION: CONNECTIVE TISSUE GROWTH FACTOR  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Spensley Horn Judas & Lubitz  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/880,031  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/167,628  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Wetherell, Jr. Ph.D., John W.  
REGISTRATION NUMBER: 31,678  
REFERENCE/DOCKET NUMBER: PD-1294  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-455-5100  
TELEFAX: 619-455-5110  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 349 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-880-031-2

Query Match 2.0%; Score 7; DB 2; Length 349;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FVLLAL 10  
|||||||  
DB 13 FVLLAL 19

RESULT 12  
US-09-054-368-2  
Sequence 2, Application US/09054368  
Patent No. 6069006  
GENERAL INFORMATION:  
APPLICANT: University of South Florida  
APPLICANT: Grotendorst, Gary R.  
APPLICANT: Bradham, Jr., Douglas M.,  
TITLE OF INVENTION: CONNECTIVE TISSUE GROWTH FACTOR  
FILE REFERENCE: 07414/003005  
CURRENT APPLICATION NUMBER: US/09/054,368  
CURRENT FILING DATE: 1998-04-02  
EARLIER APPLICATION NUMBER: 08/386,680  
EARLIER FILING DATE: 1995-02-10  
EARLIER APPLICATION NUMBER: 08/459,717  
EARLIER FILING DATE: 1995-06-02  
EARLIER APPLICATION NUMBER: 08/167,628  
EARLIER FILING DATE: 1993-12-14  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO: 2  
LENGTH: 349  
TYPE: PRT  
ORGANISM: Homo Sapiens  
US-09-054-368-2

Query Match 2.0%; Score 7; DB 3; Length 349;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FVLLAL 10  
|||||||  
DB 13 FVLLAL 19

RESULT 13  
US-09-097-179-2  
Sequence 2, Application US/09097179  
Patent No. 6149916  
GENERAL INFORMATION:  
APPLICANT: Grotendorst, Gary R.  
APPLICANT: Bradham Jr., Douglas M.,  
TITLE OF INVENTION: CONNECTIVE TISSUE GROWTH FACTOR  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Spensley Horn Judas & Lubitz  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/097,179  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/386,680

;; FILING DATE: 10-FEB-1995  
;; APPLICATION NUMBER: US/08/167,628  
;; FILING DATE:  
;; APPLICATION NUMBER: US/07/752,427  
;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Welcherell, Jr., Ph.D., John W.  
;; REGISTRATION NUMBER: 31,678  
;; REFERENCE/DOCKET NUMBER: PD-1294  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 619-455-5100  
;; TELEFAX: 619-455-5110  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 349 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-09-097-179-2

Query Match 2.0%; Score 7; DB 4; Length 349;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 FVVLAL 10  
|||||  
DB 13 FVVLAL 19

RESULT 14  
US-09-054-274-2  
;; Sequence 2, Application US/09054274  
;; Patent No. 6150101  
;; GENERAL INFORMATION:  
;; APPLICANT: University of South Florida  
;; APPLICANT: Grotendorst, Gary R.  
;; APPLICANT: Bradham, Jr., Douglas M.  
;; TITLE OF INVENTION: METHODS OF IDENTIFYING A COMPOSITION  
;; TITLE OF INVENTION: THAT ALTERS CONNECTIVE TISSUE GROWTH  
;; FILE REFERENCE: 07414/003004  
;; CURRENT APPLICATION NUMBER: US/09/054,274  
;; CURRENT FILING DATE: 1998-04-02  
;; EARLIER APPLICATION NUMBER: 08/386,680  
;; EARLIER FILING DATE: 1995-02-10  
;; EARLIER APPLICATION NUMBER: 08/459,717  
;; EARLIER FILING DATE: 1995-06-02  
;; EARLIER APPLICATION NUMBER: 08/167,628  
;; EARLIER FILING DATE: 1993-12-14  
;; NUMBER OF SEQ ID NOS: 9  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 2  
;; LENGTH: 349  
;; TYPE: PRT  
;; ORGANISM: Homo Sapiens  
US-09-054-274-2

Query Match 2.0%; Score 7; DB 4; Length 349;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 FVVLAL 10  
|||||  
DB 13 FVVLAL 19

RESULT 15  
US-09-080-715-2  
;; Sequence 2; Application US/09080715  
;; Patent No. 6150884  
;; GENERAL INFORMATION:

;; APPLICANT: Grotendorst, Gary R.  
;; APPLICANT: Bradham Jr., Douglas M.,  
;; TITLE OF INVENTION: CONNECTIVE TISSUE GROWTH FACTOR  
;; NUMBER OF SEQUENCES: 2  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Spensley Horn Judas & Lubitz  
;; STREET: 4225 Executive Square, Suite 1400  
;; CITY: La Jolla  
;; STATE: CA  
;; COUNTRY: US  
;; ZIP: 92037  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/080,715  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/167,628  
;; FILING DATE:  
;; APPLICATION NUMBER: US/07/752,427  
;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Welcherell, Jr., Ph.D., John W.  
;; REGISTRATION NUMBER: 31,678  
;; REFERENCE/DOCKET NUMBER: PD-1294  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 619-455-5100  
;; TELEFAX: 619-455-5110  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 349 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-09-080-715-2

Query Match 2.0%; Score 7; DB 4; Length 349;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 FVVLAL 10  
|||||  
DB 13 FVVLAL 19

RESULT 16  
US-09-056-704-2  
;; Sequence 2, Application US/09056704  
;; Patent No. 6232064  
;; GENERAL INFORMATION:  
;; APPLICANT: University of South Florida  
;; APPLICANT: Grotendorst, Gary R.  
;; APPLICANT: Bradham, Jr., Douglas M.  
;; TITLE OF INVENTION: METHODS OF DIAGNOSING A PATHOLOGY  
;; TITLE OF INVENTION: DISORDER ASSOCIATED BY A CELL PROLIFERATIVE  
;; TITLE OF INVENTION: TISSUE GROWTH FACTOR (Amended)  
;; FILE REFERENCE: 07414/003002  
;; CURRENT APPLICATION NUMBER: US/09/056,704  
;; CURRENT FILING DATE: 1998-03-03  
;; EARLIER APPLICATION NUMBER: 08/386,680  
;; EARLIER FILING DATE: 1995-02-10  
;; EARLIER APPLICATION NUMBER: 08/459,717  
;; EARLIER FILING DATE: 1995-06-02  
;; EARLIER APPLICATION NUMBER: 08/167,628  
;; EARLIER FILING DATE: 1993-12-14  
;; NUMBER OF SEQ ID NOS: 9  
;; SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 2  
LENGTH: 349  
TYPE: PRT  
ORGANISM: Homo Sapiens  
US-09-056-704-2

Query Match 2.0%; Score 7; DB 4; Length 349;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FVVLAL 10  
Db 13 FVVLAL 19

RESULT 17  
PCT-US96-08140-2  
Sequence 2, Application PC/TUS9608140  
GENERAL INFORMATION:  
APPLICANT: University of South Florida  
TITLE OF INVENTION: CONNECTIVE TISSUE GROWTH FACTOR  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & RICHARDSON P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/08140  
FILING DATE: 30-MAY-1996  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Halle, Ph.D., Lisa A  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07414/003W01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-678-5070  
TELEFAX: 619-678-5099  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 349 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US96-08140-2

Query Match 2.0%; Score 7; DB 5; Length 349;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FVVLAL 10  
Db 13 FVVLAL 19

RESULT 18  
US-08-969-106-6  
Sequence 6, Application US/08969106  
Patent No. 5986055  
GENERAL INFORMATION:  
APPLICANT: Yang, M.  
APPLICANT: Nandabalan, K.  
APPLICANT: Schultz, V.  
TITLE OF INVENTION: CDK2 INTERACTIONS

NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/969,106  
FILING DATE: 13-NOV-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mirock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7934-057  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 377 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-969-106-6

Query Match 2.0%; Score 7; DB 2; Length 377;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 SSLPLNS 172  
Db 252 SSLPLNS 258

RESULT 19  
US-09-054-492B-1  
Sequence 1, Application US/09054492B  
Patent No. 6218115  
GENERAL INFORMATION:  
APPLICANT: TAKESHI NAKAMURA  
TITLE OF INVENTION: HUMAN CYCLIN I AND GENES ENCODING SAME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PILLSBURY, MADISON & SUTRO  
STREET: 1100 NEW YORK AVENUE, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PATENTIN RELEASE #1.0, VERSION #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/054,492B  
FILING DATE: APRIL 3, 1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: PAUL E. WHITE, JR.  
REGISTRATION NUMBER: 32,011  
REFERENCE/DOCKET NUMBER: 7898/252159  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 671462TUSH  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 377  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-054-492B-1

Query Match 2.0%; Score 7; DB 4; Length 377;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 SSLPLNS 172  
|||||  
DB 252 SSSLPLNS 258

RESULT 20  
US-08-368-803-7  
Sequence 7, Application US/08368803  
Patent No. 5733554  
GENERAL INFORMATION:  
APPLICANT: AUDONNET, Jean-Christophe F  
APPLICANT: BUBLOT, Michel J  
APPLICANT: DARREIL, Raphael J  
APPLICANT: DUNAT, Carole V  
APPLICANT: LAPLACE, Eliane L  
APPLICANT: RIVIERE, Michel A  
TITLE OF INVENTION: Avian Herpesvirus-based live recombinant avian  
TITLE OF INVENTION: vaccine, in particular against Gumboro disease  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LARSON AND TAYLOR  
STREET: 727 SOUTH 23RD STREET  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/368,803  
FILING DATE: 05-JAN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: SARRO, Thomas  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 920-7200  
TELEFAX: (703) 892-8428  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 401 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-368-803-7

Query Match 2.0%; Score 7; DB 1; Length 401;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 KGVNKL 87  
|||||  
DB 391 KGVNKL 397

QY 81 KGVNKL 87  
|||||  
DB 391 KGVNKL 397

RESULT 21  
US-08-663-566A-5  
Sequence 5, Application US/08663566A  
Patent No. 5853733  
GENERAL INFORMATION:  
APPLICANT: Cochran, Mark D  
APPLICANT: Macdonald, Richard D  
TITLE OF INVENTION: Recombinant Herpesvirus of Turkeys  
TITLE OF INVENTION: and Uses Thereof  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John P. White  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/663,566A  
FILING DATE: June 13, 1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P  
REGISTRATION NUMBER: 28,678  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)278-0400  
TELEFAX: (212)391-0526  
TELEX: 422523  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-663-566A-5

Query Match 2.0%; Score 7; DB 2; Length 422;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 KGVNKL 87  
|||||  
DB 391 KGVNKL 397

RESULT 22  
US-08-023-610-5  
Sequence 5, Application US/08023610  
Patent No. 5928648  
GENERAL INFORMATION:  
APPLICANT: Cochran Ph.D., Mark D  
APPLICANT: Macdonald Ph.D., Richard D  
TITLE OF INVENTION: Recombinant Herpesvirus of Turkeys  
TITLE OF INVENTION: and Uses Thereof  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John P. White  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/023 610  
FILING DATE: February 26, 1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White Esq, John P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)977-9550  
TELEFAX: (212)664-0525  
TELEX: 422523  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-023-610-5

Query Match 2.0%; Score 7; DB 2; Length 422;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 KGVNKL 87  
|||||  
DB 391 KGVNKL 397

RESULT 23  
US-08-288-065A-5  
Sequence 5, Application US/08288065A  
Patent No. 5961982  
GENERAL INFORMATION:  
APPLICANT: Cochran, Mark D  
APPLICANT: Macdonald, Richard D  
TITLE OF INVENTION: Recombinant Herpesvirus of Turkeys S-  
TITLE OF INVENTION: HVT-050 and Uses Thereof  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John P. White  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/288,065A  
FILING DATE: Aug-09-94  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)278-0400  
TELEFAX: (212)391-0526  
TELEX: 422523  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-288-065A-5

Query Match 2.0%; Score 7; DB 2; Length 422;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 KGVNKL 87  
|||||  
DB 391 KGVNKL 397

RESULT 24  
US-08-362-240A-5  
Sequence 5, Application US/08362240A  
Patent No. 5965138  
GENERAL INFORMATION:  
APPLICANT: Cochran, Mark D  
APPLICANT: Junker, David  
APPLICANT: Wild, Martha A  
TITLE OF INVENTION: Recombinant Herpesvirus and Uses Thereof  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John P. White  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/362,240A  
FILING DATE: Dec-22-94  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P  
REGISTRATION NUMBER: 28,678  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)278-0400  
TELEFAX: (212)391-0526  
TELEX: 422523  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-362-240A-5

Query Match 2.0%; Score 7; DB 2; Length 422;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 KGVNKL 87  
|||||  
DB 391 KGVNKL 397

RESULT 25  
US-08-804-372A-3  
Sequence 3, Application US/08804372A  
Patent No. 6183753  
GENERAL INFORMATION:  
APPLICANT: Cochran, Mark D.  
APPLICANT: Winslow, Barbara J.  
TITLE OF INVENTION: Recombinant Chimeric Viruses and Uses  
TITLE OF INVENTION: Thereof  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York

COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,372A  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 2552/39115E  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-804-372A-3

Query Match 2.0%; Score 7; DB 4; Length 422;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 81 KGVNKL 87  
|||||  
DB 391 KGVNKL 397

RESULT 26  
PCT-US95-10245-5  
Sequence 5, Application PC/TUS9510245  
GENERAL INFORMATION:  
APPLICANT: SYNPRO CORPORATION  
TITLE OF INVENTION: Recombinant Herpesvirus of Turkeys And Uses Thereof  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John P. White  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/10245  
FILING DATE: 09-AUG-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)278-0400  
TELEFAX: (212)391-0526  
TELEX: 422523  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 422 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-10245-5

Query Match 2.0%; Score 7; DB 5; Length 422;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 81 KGVNKL 87  
|||||  
DB 391 KGVNKL 397

RESULT 27  
US-08-611-280-2  
Sequence 2, Application US/08611280  
Patent No. 5891666  
GENERAL INFORMATION:  
APPLICANT: Matsuyama, Toshifumi  
APPLICANT: Grossman, Alex  
TITLE OF INVENTION: NOVEL GENES ENCODING LSTRF POLYPEPTIDES  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Canada Inc.  
STREET: 6733 Mississauga Road, Suite 303  
CITY: Mississauga  
STATE: Ontario  
COUNTRY: Canada  
ZIP: L5N 6J8  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/611,280  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Oleski, Nancy A.  
REGISTRATION NUMBER: 34,688  
REFERENCE/DOCKET NUMBER: A-338A  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 450 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-611-280-2

Query Match 2.0%; Score 7; DB 2; Length 450;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 183 FLSLQV 189  
|||||  
DB 364 FLSLQV 370

RESULT 28  
US-09-000-145-3  
Sequence 3, Application US/09000145  
Patent No. 6169172  
GENERAL INFORMATION:  
APPLICANT: DEVAUCHELLE, Gerrard  
APPLICANT: GARNIER, Laurence  
APPLICANT: CAHOREAU, Claire  
APPLICANT: CERUTTI, Marlene  
TITLE OF INVENTION: USE OF A PROLACTIN RECEPTOR OR GROWTH HORMONE RECEPTOR  
TITLE OF INVENTION: INTRACITOPASMIC DOMAIN FOR ACHIEVING PROTEIN SECRETION  
FILE REFERENCE: 0384-0047-0XPCT  
CURRENT APPLICATION NUMBER: US/09/000,145

;; CURRENT FILING DATE: 1998-03-16  
;; EARLIER APPLICATION NUMBER: PCT/EP96/01237  
;; EARLIER FILING DATE: 1996-08-02  
;; EARLIER APPLICATION NUMBER: PR 95/09420  
;; EARLIER FILING DATE: 1995-08-02  
;; NUMBER OF SEQ ID NOS: 6  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 3  
;; TYPE: PRT  
;; ORGANISM: Oryctolagus cuniculus  
US-09-000-145-3

Query Match 2.0%; Score 7; DB 4; Length 620;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 LKSPGSV 28  
|||||  
Db 58 LKSPGSV 64

RESULT 29  
US-09-413-814-107  
; Sequence 107, Application US/09413814  
; Patent No. 6225064  
; GENERAL INFORMATION:  
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
; APPLICANT: Bristol-Myers Squibb, Co.  
; APPLICANT: Beyer, Stefan  
; APPLICANT: Bloeker, Helmut  
; APPLICANT: Brandt, Petra  
; APPLICANT: Cino, Paul M  
; APPLICANT: Dougherty, Brian A  
; APPLICANT: Goldberg, Steven L  
; APPLICANT: Hoffe, Gerhard  
; APPLICANT: Mueller, Joachim  
; APPLICANT: Reichenbach, Hans  
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or  
; TITLE OF INVENTION: heteropolyketide compounds  
; FILE REFERENCE: PCT/US 99/23535  
; CURRENT APPLICATION NUMBER: US/09/413,814  
; CURRENT FILING DATE: 1999-10-07  
; EARLIER APPLICATION NUMBER: DE 198 46 493.2  
; EARLIER FILING DATE: 1998-10-09  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 107  
; LENGTH: 841  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-413-814-107

Query Match 2.0%; Score 7; DB 4; Length 841;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LALVAGV 14  
|||||  
Db 227 LALVAGV 233

RESULT 30  
US-09-413-814-28  
; Sequence 28, Application US/09413814  
; Patent No. 6225064  
; GENERAL INFORMATION:  
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
; APPLICANT: Bristol-Myers Squibb, Co.  
; APPLICANT: Beyer, Stefan  
; APPLICANT: Bloeker, Helmut

;; APPLICANT: Brandt, Petra  
;; APPLICANT: Cino, Paul M  
;; APPLICANT: Dougherty, Brian A  
;; APPLICANT: Goldberg, Steven L  
;; APPLICANT: Hoffe, Gerhard  
;; APPLICANT: Mueller, Joachim  
;; APPLICANT: Reichenbach, Hans  
;; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or  
;; TITLE OF INVENTION: heteropolyketide compounds  
;; FILE REFERENCE: PCT/US 99/23535  
;; CURRENT APPLICATION NUMBER: US/09/413,814  
;; CURRENT FILING DATE: 1999-10-07  
;; EARLIER APPLICATION NUMBER: DE 198 46 493.2  
;; EARLIER FILING DATE: 1998-10-09  
;; NUMBER OF SEQ ID NOS: 107  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 28  
;; LENGTH: 2618  
;; TYPE: PRT  
;; ORGANISM: Sorangium cellulosum  
US-09-413-814-28

Query Match 2.0%; Score 7; DB 4; Length 2618;  
Best Local Similarity 100.0%; Pred. No. 6.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 ELAIGDE 221  
|||||  
Db 281 ELAIGDE 287

RESULT 31  
US-08-159-339A-917  
; Sequence 917, Application US/08159339A  
; Patent No. 6037135  
; GENERAL INFORMATION:  
; APPLICANT: Kubo, Ralph T.  
; APPLICANT: Grey, Howard M.  
; APPLICANT: Sette, Alessandro  
; APPLICANT: Celis, Esben  
; TITLE OF INVENTION: HLA Binding peptides and their  
; TITLE OF INVENTION: uses  
; NUMBER OF SEQUENCES: 1254  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/159,339A  
; FILING DATE: 29-NOV-1993  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/926,666  
; FILING DATE: 07-AUG-1992  
; APPLICATION NUMBER: US 08/027,746  
; FILING DATE: 05-MAR-1993  
; APPLICATION NUMBER: US 08/103,396  
; FILING DATE: 06-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weber, Ellen Lauver  
; REGISTRATION NUMBER: 32,762  
; REFERENCE/DOCKET NUMBER: 018623-005030US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200



TELEFAX: (415) 576-0300  
TELEX:  
INFORMATION FOR SEQ ID NO: 917:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-159-339A-917

Query Match 1.7%; Score 6; DB 3; Length 9;  
Best Local Similarity 100.0%; Pred. No. 14e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 280 ILEAKQ 285  
DB 1 ILEAKQ 6

RESULT 32  
US-07-958-903A-18  
Sequence 18, Application US/07958903A  
Patent No. 5652214  
GENERAL INFORMATION:  
APPLICANT: Lewis, Michael E.  
APPLICANT: Kauer, James C.  
APPLICANT: Smith, Kevin R.  
APPLICANT: Callison, Kathleen V.  
APPLICANT: Baldino, Frank  
APPLICANT: Neff, Nicola  
APPLICANT: Iqbal, Mohamed  
TITLE OF INVENTION: TREATING DISORDERS BY APPLICATION  
TITLE OF INVENTION: OF INSULIN-LIKE GROWTH FACTORS AND  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESS: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 555X  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: Wordperfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/958, 903A  
FILING DATE: October 7, 1992  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/361,595  
FILING DATE: June 5, 1989  
APPLICATION NUMBER: 07/534,139  
FILING DATE: June 5, 1990  
APPLICATION NUMBER: 07/869,913  
FILING DATE: April 15, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 02655/003004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 13  
TYPE: amino acid  
STRANDEDNESS:

TOPOLOGY: linear  
US-07-958-903A-18

Query Match 1.7%; Score 6; DB 1; Length 13;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 240 LVPDLQ 245  
DB 5 LVPDLQ 10

RESULT 33  
US-08-462-018-18  
Sequence 18, Application US/08462018  
Patent No. 5703045  
GENERAL INFORMATION:  
APPLICANT: Lewis, Michael E.  
APPLICANT: Kauer, James C.  
APPLICANT: Smith, Kevin R.  
APPLICANT: Callison, Kathleen V.  
APPLICANT: Baldino, Frank  
APPLICANT: Neff, Nicola  
APPLICANT: Iqbal, Mohamed  
TITLE OF INVENTION: TREATING DISORDERS BY APPLICATION  
TITLE OF INVENTION: OF INSULIN-LIKE GROWTH FACTORS AND  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESS: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 555X  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: Wordperfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/462,018  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/958, 903  
FILING DATE: October 7, 1992  
APPLICATION NUMBER: 07/361,595  
FILING DATE: June 5, 1989  
APPLICATION NUMBER: 07/534,139  
FILING DATE: June 5, 1990  
APPLICATION NUMBER: 07/869, 913  
FILING DATE: April 15, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 02655/003005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 13  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-462-018-18  
Query Match 1.7%; Score 6; DB 1; Length 13;  
Best Local Similarity 100.0%; Pred. No. 45;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 240 LVDAIQ 245  
Db 5 LVDAIQ 10

RESULT 34  
US-08-823-245-18  
; Sequence 18, Application US/08823245  
; Patent No. 5776897  
; GENERAL INFORMATION:  
; APPLICANT: Lewis, Michael  
; APPLICANT: Kauer, James C.  
; APPLICANT: Smith, Kevin R.  
; APPLICANT: Callison, Kathleen V.  
; APPLICANT: Baldino, Frank  
; APPLICANT: Neft, Nicola  
; APPLICANT: Iqbal, Mohamed  
; TITLE OF INVENTION: TREATING DISORDERS BY  
; TITLE OF INVENTION: APPLICATION  
; TITLE OF INVENTION: OF INSULIN-LIKE GROWTH  
; TITLE OF INVENTION: FACTORS AND  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Fish & Richardson  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
; COMPUTER: IBM PS/2 Model 502 or  
; OPERATING SYSTEM: MS-DOS (Version 5.0)  
; SOFTWARE: Wordperfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/823,245  
; FILING DATE: March 24, 1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/361,595  
; FILING DATE: June 6, 1989  
; APPLICATION NUMBER: 07/534,139  
; FILING DATE: June 5, 1990  
; APPLICATION NUMBER: 07/869,913  
; FILING DATE: April 15, 1992  
; APPLICATION NUMBER: 07/938,903  
; FILING DATE: October 7, 1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cireson, Gary L.  
; REGISTRATION NUMBER: 34,310  
; REFERENCE/DOCKET NUMBER: 02655/003008  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 13  
; TYPE: amino acid  
; STRANDEDNESS: N/A  
; TOPOLOGY: N/A  
US-08-823-245-18

Query Match 1.7%; Score 6; DB 1; Length 13;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 240 LVDAIQ 245  
Db 5 LVDAIQ 10

RESULT 35  
PCT-US92-09443A-18  
; Sequence 18, Application PC/7US9209443A  
; GENERAL INFORMATION:  
; APPLICANT: Bozyczko-Coyne, Donna  
; APPLICANT: Neft, Nicola  
; APPLICANT: Lewis, Michael E.  
; APPLICANT: Iqbal, Mohamed  
; TITLE OF INVENTION: TREATING RETINAL NEURONAL  
; TITLE OF INVENTION: DISORDERS BY THE APPLICATION OF  
; TITLE OF INVENTION: INSULIN-LIKE GROWTH FACTORS AND  
; NUMBER OF SEQUENCES: 79  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Fish & Richardson  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
; COMPUTER: IBM PS/2 Model 502 or 555X  
; OPERATING SYSTEM: MS-DOS (Version 5.0)  
; SOFTWARE: Wordperfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/09443A  
; FILING DATE: 19921103  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/790,690  
; FILING DATE: November 8, 1991  
; APPLICATION NUMBER: 07/963,329  
; FILING DATE: October 15, 1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clark, Paul T.  
; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 02655/012W02  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 13  
; TYPE: AMINO ACID  
; STRANDEDNESS: N/A  
; TOPOLOGY: N/A  
PCT-US92-09443A-18

Query Match 1.7%; Score 6; DB 5; Length 13;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 240 LVDAIQ 245  
Db 5 LVDAIQ 10

RESULT 36  
US-08-350-260A-46  
; Sequence 46, Application US/08350260A  
; Patent No. 5962255  
; GENERAL INFORMATION:  
; APPLICANT: Winter, Gregory Paul  
; APPLICANT: Griffiths, Andrew David  
; APPLICANT: Williams, Samuel Cameron

```

APPLICANT: Waterhouse, Peter
APPLICANT: Missim, Ahuva
APPLICANT: Johnson, Kevin Stuart
APPLICANT: Smith, Andrew John Hammond
TITLE OF INVENTION: Methods for producing members of specific
TITLE OF INVENTION: binding pairs
NUMBER OF SEQUENCES: 602
CORRESPONDENCE ADDRESS:
ADDRESSEE: David W. Clough
STREET: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
ZIP: 60606-6402
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/350,260A
FILING DATE: 05-DEC-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9110549.4
FILING DATE: 15-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9206318.9
FILING DATE: 24-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB91/01134
FILING DATE: 10-JUL-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB92/00883
FILING DATE: 15-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/00605
FILING DATE: 24-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/150,002
FILING DATE: 31-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/307,619
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 28111/32372
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-350-260A-46

Query Match 1.7%; Score 6; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 323 ITSYNI 328
Db 1 ITSYNI 6

RESULT 37
US-08-350-260A-48
; Sequence 48, Application US/08350260A
; Patent No. 5962255

```

```

GENERAL INFORMATION:
APPLICANT: Winter, Gregory Paul
APPLICANT: Griffiths, Andrew David
APPLICANT: Williams, Samuel Cameron
APPLICANT: Waterhouse, Peter
APPLICANT: Missim, Ahuva
APPLICANT: Johnson, Kevin Stuart
APPLICANT: Smith, Andrew John Hammond
TITLE OF INVENTION: Methods for producing members of specific
TITLE OF INVENTION: binding pairs
NUMBER OF SEQUENCES: 602
CORRESPONDENCE ADDRESS:
ADDRESSEE: David W. Clough
STREET: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
ZIP: 60606-6402
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/350,260A
FILING DATE: 05-DEC-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9110549.4
FILING DATE: 15-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9206318.9
FILING DATE: 24-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB91/01134
FILING DATE: 10-JUL-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB92/00883
FILING DATE: 15-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/00605
FILING DATE: 24-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/150,002
FILING DATE: 31-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/307,619
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 28111/32372
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-350-260A-48

Query Match 1.7%; Score 6; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 323 ITSYNI 328
Db 1 ITSYNI 6

```

RESULT 38  
US-08-637-759B-65  
; Sequence 65, Application US/08637759B  
; Patent No. 5876331  
; GENERAL INFORMATION:  
; APPLICANT: David William Holden  
; TITLE OF INVENTION: Identification of Genes  
; NUMBER OF SEQUENCES: 501  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Patrea L. Pabst  
; STREET: 2800 One Atlantic Center  
; STREET: 1201 West Peachtree Street  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30309-3450  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/637,759B  
; FILING DATE: 03-MAY-1996  
; CLASSIFICATION: 435  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB95/02875  
; FILING DATE: 11-DEC-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pabst, Patrea L.  
; REGISTRATION NUMBER: 31,284  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404) 873-8794  
; TELEFAX: (404) 873-8794  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 17 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
; US-08-637-759B-65

Query Match 1.7%; Score 6; DB 2; Length 17;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 ISYPLE 99  
DB 5 ISYPLE 10

RESULT 39  
US-08-871-355A-65  
; Sequence 65, Application US/08871355A  
; Patent No. 6015669  
; GENERAL INFORMATION:  
; APPLICANT: David William Holden  
; TITLE OF INVENTION: Identification of Genes  
; NUMBER OF SEQUENCES: 501  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Patrea L. Pabst  
; STREET: 2800 One Atlantic Center  
; STREET: 1201 West Peachtree Street  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30309-3450  
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/871,355A  
FILING DATE: 09-JUN-1997  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB95/02875  
FILING DATE: 11-DEC-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: RPKS 101 CON  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 873-8794  
TELEFAX: (404) 873-8794  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: NO  
US-08-871-355A-65

Query Match 1.7%; Score 6; DB 3; Length 17;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 ISYPLE 99  
DB 5 ISYPLE 10

RESULT 40  
US-08-484-530-34  
; Sequence 34, Application US/08484530  
; Patent No. 5846740  
; GENERAL INFORMATION:  
; APPLICANT: Todin, Allan J  
; APPLICANT: Eriander, Mark G  
; APPLICANT: Kautman, Daniel L.  
; TITLE OF INVENTION: Cloned Glutamic Acid Decarboxylase  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/484,530  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Tregerlin, Richard F.  
; REGISTRATION NUMBER: 31,801  
; REFERENCE/DOCKET NUMBER: A-60780-8/RET/MTK  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-781-1989  
; TELEFAX: 415-398-3249  
; INFORMATION FOR SEQ ID NO: 34:

SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-484-530-34

Query Match 1.7%; Score 6; DB 2; Length 20;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 280 ILEAKQ 285  
DB 15 ILEAKQ 20

RESULT 41  
US-08-827-618A-34  
; Sequence 34, Application US/08827618A  
; Patent No. 5998366  
; GENERAL INFORMATION:  
; APPLICANT: Tobin, Allan J  
; APPLICANT: Erlander, Mark G  
; APPLICANT: Kaufman, Daniel L.  
; TITLE OF INVENTION: Cloned Glutamic Acid Decarboxylase  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: US  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/827,618A  
; FILING DATE: 09-APR-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,725  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/716,909  
; FILING DATE: 18-JUN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/586,536  
; FILING DATE: 21-SEP-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Trecartin, Richard F.  
; REGISTRATION NUMBER: 31,801  
; REFERENCE/DOCKET NUMBER: A-60780-12/RTT/MTK  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-781-1989  
; TELEFAX: 415-398-3249  
; INFORMATION FOR SEQ ID NO: 34:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-827-618A-34

Query Match 1.7%; Score 6; DB 2; Length 20;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 280 ILEAKQ 285  
DB 15 ILEAKQ 20

RESULT 42  
US-08-483-952A-34  
; Sequence 34, Application US/08483952A  
; Patent No. 6011139  
; GENERAL INFORMATION:  
; APPLICANT: Tobin, Allan J  
; APPLICANT: Erlander, Mark G  
; APPLICANT: Kaufman, Daniel L.  
; TITLE OF INVENTION: Cloned Glutamic Acid Decarboxylase  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: US  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/483,952A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/123,859  
; FILING DATE: 17-SEP-1993  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/716,909  
; FILING DATE: 18-JUN-1991  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/586,536  
; FILING DATE: 21-SEP-1990  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Trecartin, Richard F.  
; REGISTRATION NUMBER: 31,801  
; REFERENCE/DOCKET NUMBER: A-60780-7/RTT/MTK  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-781-1989  
; TELEFAX: 415-398-3249  
; INFORMATION FOR SEQ ID NO: 34:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-483-952A-34

Query Match 1.7%; Score 6; DB 3; Length 20;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 280 ILEAKQ 285  
DB 15 ILEAKQ 20

RESULT 43  
US-08-602-999A-143  
; Sequence 143, Application US/08602999A

Patent No. 6184205  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: KAY, Brian K.  
APPLICANT: THORN, Judith M.  
APPLICANT: OULLIAM, Lawrence A.  
APPLICANT: DER, Channing J.  
APPLICANT: FOWLER, Dana M.  
APPLICANT: RIDER, James E.  
TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
NUMBER OF SEQUENCES: 467  
CORRESPONDENCE ADDRESS:  
ADDRESS: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,999A  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 143:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-602-999A-143

Query Match 1.7%; Score 6; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 LNSLSR 175  
DB 15 LNSLSR 20

RESULT 44  
PCT-US92-09443A-73  
Sequence 73, Application PC/TUS9209443A  
GENERAL INFORMATION:  
APPLICANT: Bozyczko-Coyne, Donna  
APPLICANT: Neff, Nicola  
APPLICANT: Lewis, Michael E.  
APPLICANT: Iqbal, Mohamed  
TITLE OF INVENTION: TREATING RETINAL NEURONAL  
TITLE OF INVENTION: DISORDERS BY THE APPLICATION OF  
TITLE OF INVENTION: INSULIN-LIKE GROWTH FACTORS AND  
TITLE OF INVENTION: ANALOGS  
NUMBER OF SEQUENCES: 79  
CORRESPONDENCE ADDRESS:  
ADDRESS: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.

ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
COMPUTER: IBM PS/2 Model 50z or 555X  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: Wordperfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/09443A  
FILING DATE: 19921103  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/790,690  
FILING DATE: November 8, 1991  
APPLICATION NUMBER: 07/963,329  
FILING DATE: October 15, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 02655/012M02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 73:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 24  
TYPE: AMINO ACID  
STRANDEDNESS: N/A  
TOPOLOGY: N/A  
PCT-US92-09443A-73

Query Match 1.7%; Score 6; DB 5; Length 24;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 240 LVDAIQ 245  
DB 4 LVDAIQ 9

RESULT 45  
US-08-090-036-2  
Sequence 2, Application US/08090036  
Patent No. 5474922  
GENERAL INFORMATION:  
APPLICANT: Doreich, Kurt  
APPLICANT: Dalboge, Henrik  
APPLICANT: Mikkelsen, Jan M.  
APPLICANT: Mischler, Marcel  
APPLICANT: Christensen, Flemming M.  
TITLE OF INVENTION: Beta-1,4-GALACTANASE AND A DNA SEQUENCE  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESS: No. 54749220 No. 54749220disk of No. 5474922th America, Inc.  
STREET: 405 Lexington Avenue, 62nd Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6201  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/090,036  
FILING DATE: 15-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DK92/00037  
FILING DATE: 06-FEB-1992  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: BE 91610007.6  
FILING DATE: 06-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Lowney Dr., Karen A.  
REGISTRATION NUMBER: 31,274  
REFERENCE/DOCKET NUMBER: 3564.204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-867-0298  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 28 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Aspergillus niger  
US-08-090-036-2

Query Match 1.7%; Score 6; DB 1; Length 28;  
Best Local Similarity 100.0%; Pred. No. 90;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 192 DISSL 197  
Db 8 DISSL 13

RESULT 46  
US-08-435-252-1  
Sequence 1, Application US/08435252  
Patent No. 5622932  
GENERAL INFORMATION:  
APPLICANT: Dimatchi, Richard D.  
APPLICANT: Li, Fan  
APPLICANT: Long, Harlan B.  
TITLE OF INVENTION: IGF-1 Superagonists  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center/Patent Division  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: U.S.A.  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/435,252  
FILING DATE: 05-MAY-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Maciak, Ronald D.  
REGISTRATION NUMBER: 35,262  
REFERENCE/DOCKET NUMBER: X8030  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-1664  
TELEFAX: 317-277-1917  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 29 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-435-252-1

Query Match 1.7%; Score 6; DB 1; Length 29;  
Best Local Similarity 100.0%; Pred. No. 92;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 240 LVDAQ 245  
Db 10 LVDAQ 15

RESULT 47  
PCT-US96-01314-68  
Sequence 68, Application PC/TUS9601314  
GENERAL INFORMATION:  
APPLICANT: M. Amin Arnaout  
TITLE OF INVENTION: METHODS FOR IDENTIFYING INTEGRIN  
NUMBER OF SEQUENCES: 78  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: Wordperfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/01314  
FILING DATE: 30-JAN-96  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/380,167  
FILING DATE: 30-JAN-95  
ATTORNEY/AGENT INFORMATION:  
NAME: John W. Freeman  
REGISTRATION NUMBER: 29,066  
REFERENCE/DOCKET NUMBER: 00786/267001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 68:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 29 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US96-01314-68

Query Match 1.7%; Score 6; DB 5; Length 29;  
Best Local Similarity 100.0%; Pred. No. 92;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 236 ASKILV 241  
Db 17 ASKILV 22

RESULT 48  
US-08-066-325-72  
Sequence 72, Application US/08066325  
Patent No. 5667967  
GENERAL INFORMATION:  
APPLICANT: Steinman, Lawrence  
APPLICANT: Oksenberg, Jorge  
APPLICANT: Bernard, Claude  
TITLE OF INVENTION: T-CELL RECEPTOR VARIABLE TRANSCRIPTS AS DISEASE RELATED MAR

NUMBER OF SEQUENCES: 157  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/066,325  
FILING DATE: 21-MAY-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5667967endburg Ph.D., Carol  
REGISTRATION NUMBER: 39,317  
REFERENCE/DOCKET NUMBER: 690068.408C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 72:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-066-325-72

Query Match  
Best Local Similarity 1.7%; Score 6; DB 1; Length 30;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 213 SLELAG 218  
|||||  
DB 5 SLELAG 10

RESULT 49  
US-08-090-036-1  
Sequence 1, Application US/08090036  
Patent No. 5474922  
GENERAL INFORMATION:  
APPLICANT: Doreich, Kurt  
APPLICANT: Dalboge, Henrik  
APPLICANT: Mikkelsen, Jan M.  
APPLICANT: Mischler, Marcel  
TITLE OF INVENTION: Beta-1,4-GALACTANASE AND A DNA SEQUENCE  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 5474922 No. 5474922disk of No. 5474922th America, Inc.  
STREET: 405 Lexington Avenue, 62nd Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6201  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/090,036  
FILING DATE: 15-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DK92/00037

FILING DATE: 06-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: BE 91610007.6  
FILING DATE: 06-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Lowney Dr., Karen A.  
REGISTRATION NUMBER: 31,274  
REFERENCE/DOCKET NUMBER: 3564.204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-867-0298  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 31 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Aspergillus ncutatus  
STRAIN: CBS 101.43  
US-08-090-036-1

Query Match  
Best Local Similarity 1.7%; Score 6; DB 1; Length 31;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 192 DISSLL 197  
|||||  
DB 8 DISSLL 13

RESULT 50  
US-08-190-802A-247  
Sequence 247, Application US/08190802A  
Patent No. 5519003  
GENERAL INFORMATION:  
APPLICANT: Mochly-Rosen, Daria  
APPLICANT: Ron, Dorit  
TITLE OF INVENTION: WD-40 - Derived Peptides and Uses  
NUMBER OF SEQUENCES: 265  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Denlinger & Associates  
STREET: P.O. Box 60850  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306-0850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/190,802A  
FILING DATE: 01-FEB-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Fabian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 8600-0139  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 247:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 31 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown





•  
•  
•

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 17, 2001, 16:44:18 ; Search time 24.67 Seconds  
(without alignments)  
1877.047 Million cell updates/sec

Title: US-09-284-320-6

Perfect score: 350  
Sequence: 1 MAVVVTLLALVAGVIGNEFS.....MDPGYDITITMNTKTRMD 350

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 425026 seqs, 132305027 residues

Word size: 5

Total number of hits satisfying chosen parameters: 35138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database:

1: SPTRMBL16.\*  
2: sp\_archaea.\*  
3: sp\_bacteria.\*  
4: sp\_fungi.\*  
5: sp\_human.\*  
6: sp\_invertebrate.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_ricent.\*  
12: sp\_unclassified.\*  
13: sp\_vertebrate.\*  
14: sp\_virus.\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	284	81.1	350	4	09H288
2	264	75.4	330	4	09G889
3	9	2.6	343	2	09G889
4	8	2.3	55	2	09G889
5	8	2.3	156	1	09H276
6	8	2.3	239	4	09H276
7	8	2.3	320	5	09H276
8	8	2.3	324	5	09H276
9	8	2.3	340	5	09H276
10	8	2.3	346	4	09H276
11	8	2.3	383	4	09H276
12	8	2.3	438	5	09H276
13	8	2.3	468	10	09H276
14	8	2.3	486	10	09H276
15	8	2.3	610	10	09H276
16	8	2.3	656	10	09H276
17	8	2.3	695	8	09H276
18	8	2.3	890	5	09H276
19	8	2.3	1413	10	09H276
20	1491	10	09H288	8	09H288
21	1501	10	09H288	8	09H288
22	1715	11	09H288	8	09H288
23	1728	11	09H288	8	09H288
24	5149	2	09H288	8	09H288
25	67	5	09H288	8	09H288
26	73	8	09H288	8	09H288
27	74	5	09H288	8	09H288
28	126	4	09H288	8	09H288
29	136	4	09H288	8	09H288
30	137	2	09H288	8	09H288
31	137	2	09H288	8	09H288
32	141	5	09H288	8	09H288
33	148	1	09H288	8	09H288
34	160	8	09H288	8	09H288
35	168	2	09H288	8	09H288
36	172	10	09H288	8	09H288
37	174	7	09H288	8	09H288
38	179	1	09H288	8	09H288
39	184	1	09H288	8	09H288
40	184	1	09H288	8	09H288
41	195	9	09H288	8	09H288
42	196	5	09H288	8	09H288
43	202	2	09H288	8	09H288
44	203	10	09H288	8	09H288
45	213	2	09H288	8	09H288
46	213	14	09H288	8	09H288
47	220	10	09H288	8	09H288
48	224	2	09H288	8	09H288
49	225	5	09H288	8	09H288
50	228	2	09H288	8	09H288
51	239	1	09H288	8	09H288
52	240	3	09H288	8	09H288
53	251	9	09H288	8	09H288
54	254	2	09H288	8	09H288
55	260	2	09H288	8	09H288
56	264	10	09H288	8	09H288
57	266	3	09H288	8	09H288
58	267	2	09H288	8	09H288
59	271	2	09H288	8	09H288
60	274	2	09H288	8	09H288
61	274	2	09H288	8	09H288
62	279	1	09H288	8	09H288
63	281	10	09H288	8	09H288
64	282	1	09H288	8	09H288
65	291	2	09H288	8	09H288
66	295	1	09H288	8	09H288
67	299	2	09H288	8	09H288
68	309	13	09H288	8	09H288
69	310	2	09H288	8	09H288
70	319	10	09H288	8	09H288
71	321	2	09H288	8	09H288
72	324	13	09H288	8	09H288
73	326	2	09H288	8	09H288
74	327	2	09H288	8	09H288
75	334	2	09H288	8	09H288
76	336	2	09H288	8	09H288
77	338	1	09H288	8	09H288
78	341	6	09H288	8	09H288
79	349	2	09H288	8	09H288
80	353	3	09H288	8	09H288
81	364	3	09H288	8	09H288
82	373	2	09H288	8	09H288
83	374	6	09H288	8	09H288
84	380	5	09H288	8	09H288
85	385	5	09H288	8	09H288
86	388	2	09H288	8	09H288
87	388	5	09H288	8	09H288
88	388	6	09H288	8	09H288
89	394	1	09H288	8	09H288
90	395	2	09H288	8	09H288
91	396	5	09H288	8	09H288
92	397	7	09H288	8	09H288
93	09H288	8	09H288	8	09H288
94	09H288	8	09H288	8	09H288
95	09H288	8	09H288	8	09H288
96	09H288	8	09H288	8	09H288
97	09H288	8	09H288	8	09H288
98	09H288	8	09H288	8	09H288
99	09H288	8	09H288	8	09H288
100	09H288	8	09H288	8	09H288

93	7	2.0	398	7	031600	031600 gallus gall	166	7	2.0	1415	4	09H7M7
94	7	2.0	400	2	P73816	P73816 synecocyst	167	7	2.0	1530	2	051583
95	7	2.0	401	1	029317	029317 archaeoglob	168	7	2.0	1596	14	084125
96	7	2.0	401	2	09HT26	09HT26 pseudomonas	169	7	2.0	1612	14	010447
97	7	2.0	401	7	031407	031407 gallus gall	170	7	2.0	1670	14	019167
98	7	2.0	408	5	09YDL3	09YDL3 drosophila	171	7	2.0	1762	2	030480
99	7	2.0	414	2	069474	069474 mycobacteri	172	7	2.0	1895	14	090A33
100	7	2.0	414	10	09LMO5	09LMO5 physcomitri	173	7	2.0	1901	14	09DHN8
101	7	2.0	422	14	09D6U9	09D6U9 meleagrid h	174	7	2.0	2076	5	001485
102	7	2.0	430	2	09KXY2	09KXY2 streptomyce	175	7	2.0	2100	2	030481
103	7	2.0	436	2	056277	056277 thioacillu	176	7	2.0	2340	11	060705
104	7	2.0	439	2	033679	033679 rhizobium m	177	7	2.0	2340	11	064736
105	7	2.0	439	10	041272	041272 arabidopsis	178	7	2.0	2434	11	09ESR9
106	7	2.0	439	10	09LHMI	09LHMI arabidopsis	179	7	2.0	2436	4	09HC28
107	7	2.0	444	11	09QZK7	09QZK7 mus musculu	180	7	2.0	2564	5	09U755
108	7	2.0	444	3	09QZK7	09QZK7 neurospora	181	7	2.0	2564	5	09K677
109	7	2.0	454	5	000896	000896 physarum po	182	7	1.7	26	4	09NP34
110	7	2.0	502	13	09DD58	09DD58 brachydanio	183	7	1.7	30	2	09RER6
111	7	2.0	506	2	09K003	09K003 neisseria m	184	7	1.7	35	2	09RAS7
112	7	2.0	506	2	09JVT2	09JVT2 neisseria m	185	7	1.7	43	11	09WUY4
113	7	2.0	507	5	09I776	09I776 drosophila	186	7	1.7	44	3	09UWU0
114	7	2.0	515	2	09XAI9	09XAI9 streptomyce	187	7	1.7	48	4	09UWU0
115	7	2.0	540	2	02RK15	02RK15 streptomyce	188	7	1.7	48	4	09UDC0
116	7	2.0	560	2	0923X9	0923X9 pseudomonas	189	7	1.7	52	2	09P015
117	7	2.0	563	6	09XSB8	09XSB8 canis famil	190	7	1.7	59	4	09UDB9
118	7	2.0	563	11	09EQV6	09EQV6 rattus norv	191	7	1.7	60	6	09GMO6
119	7	2.0	571	1	09YB72	09YB72 aeropyrum p	192	7	1.7	63	13	09PRT5
120	7	2.0	576	5	09W4F7	09W4F7 drosophila	193	7	1.7	65	4	012806
121	7	2.0	586	5	09N429	09N429 caenorhabdi	194	7	1.7	67	2	09X333
122	7	2.0	600	4	09UPF3	09UPF3 homo sapien	195	7	1.7	67	11	090970
123	7	2.0	607	4	09H9P9	09H9P9 homo sapien	196	7	1.7	68	2	09X334
124	7	2.0	626	5	09NXC4	09NXC4 homo sapien	197	7	1.7	72	10	096314
125	7	2.0	631	5	001375	001375 teratohabd	198	7	1.7	76	2	049681
126	7	2.0	633	2	09KR21	09KR21 vibrio chol	199	7	1.7	76	2	047103
127	7	2.0	646	5	017913	017913 caenorhabdi	200	7	1.7	76	2	049681
128	7	2.0	647	4	09H6T8	09H6T8 homo sapien	201	7	1.7	79	2	09GMD7
129	7	2.0	648	5	020191	020191 caenorhabdi	202	7	1.7	79	2	050974
130	7	2.0	660	5	09VFW5	09VFW5 drosophila	203	7	1.7	80	4	09UDB8
131	7	2.0	662	10	09FMS4	09FMS4 chlamydomon	204	7	1.7	80	6	029183
132	7	2.0	664	3	014120	014120 schizosach	205	7	1.7	82	2	09KEX7
133	7	2.0	689	10	09FWM0	09FWM0 arabidopsis	206	7	1.7	82	5	018908
134	7	2.0	728	10	09SCV0	09SCV0 arabidopsis	207	7	1.7	85	7	095596
135	7	2.0	729	10	09S215	09S215 arabidopsis	208	7	1.7	85	7	09VME7
136	7	2.0	744	2	09ZCD4	09ZCD4 rickettsia f	209	7	1.7	87	14	086539
137	7	2.0	761	5	09N2P1	09N2P1 citrithidia f	210	7	1.7	88	2	085938
138	7	2.0	770	4	075074	075074 homo sapien	211	7	1.7	88	2	09PC98
139	7	2.0	770	11	088204	088204 taltus norv	212	7	1.7	89	2	09RRN0
140	7	2.0	788	3	007381	007381 saccharomyc	213	7	1.7	91	2	051800
141	7	2.0	808	1	058937	058937 methanococ	214	7	1.7	91	10	09ZTX1
142	7	2.0	816	5	09N9P7	09N9P7 leishmania	215	7	1.7	91	10	09XHZ4
143	7	2.0	830	5	09NFC07	09NFC07 caenorhabdi	216	7	1.7	91	10	09GTR0
144	7	2.0	847	10	09STES	09STES arabidopsis	217	7	1.7	93	10	09SFR3
145	7	2.0	863	10	09SKF9	09SKF9 arabidopsis	218	7	1.7	94	2	087046
146	7	2.0	886	4	060341	060341 homo sapien	219	7	1.7	96	2	032559
147	7	2.0	909	5	09N618	09N618 caenorhabdi	220	7	1.7	96	2	09JRK8
148	7	2.0	950	4	09NMH3	09NMH3 homo sapien	221	7	1.7	96	2	09EYB7
149	7	2.0	961	2	049551	049551 mycoplasma	222	7	1.7	96	9	048423
150	7	2.0	974	2	09Z6S8	09Z6S8 chlamydia p	223	7	1.7	96	9	09XR80
151	7	2.0	995	3	09J580	09J580 chlamydia p	224	7	1.7	97	2	09RUV1
152	7	2.0	1011	3	09W2L7	09W2L7 drosophila	225	7	1.7	97	2	09NUN3
153	7	2.0	1020	2	09P944	09P944 pneumocysti	226	7	1.7	98	4	09H4C2
154	7	2.0	1020	2	09ZAS6	09ZAS6 salmonella	227	7	1.7	98	14	09Q0U1
155	7	2.0	1111	14	090348	090348 odontogloss	228	7	1.7	99	5	090AD8
156	7	2.0	1112	14	084124	084124 odontogloss	229	7	1.7	100	5	09V489
157	7	2.0	1112	14	084134	084134 odontogloss	230	7	1.7	100	5	09UAE3
158	7	2.0	1113	14	039641	039641 odontogloss	231	7	1.7	100	5	09UAD9
159	7	2.0	1121	5	09P680	09P680 xyella fas	232	7	1.7	100	5	09UAD5
160	7	2.0	1138	5	09NNA9	09NNA9 caenorhabdi	233	7	1.7	100	5	09UAD5
161	7	2.0	1156	10	09SJC9	09SJC9 arabidopsis	234	7	1.7	100	5	09UAD5
162	7	2.0	1166	14	09ONC4	09ONC4 cucumber gr	235	7	1.7	101	3	032897
163	7	2.0	1171	2	09HMB5	09HMB5 pseudomonas	236	7	1.7	101	13	093442
164	7	2.0	1213	5	09VCK7	09VCK7 drosophila	237	7	1.7	101	14	09YRD3
165	7	2.0	1304	5	015765	015765 tetrahymena	238	7	1.7	101	14	09YRD2

09H7M7 homo sapien  
 051583 plectonema  
 084125 odontogloss  
 010447 odontogloss  
 09167 cucumber gr  
 030480 streptomyce  
 09QAY3 ybba monkey  
 09DHN8 ybba like d  
 001485 caenorhabdi  
 030481 streptomyce  
 060705 mus musculu  
 064736 mus musculu  
 09H28 homo sapien  
 09U755 toxoplasma  
 09K97 aeromonas h  
 09NP34 homo sapien  
 09ER6 enterobacte  
 091457 rhodobacter  
 09WUY4 rattus norv  
 09UWU0 tylospora f  
 P92073 eubadra her  
 09UDC0 homo sapien  
 09P015 xyella fas  
 09UDB9 homo sapien  
 09GMO6 macaca fasc  
 09PRT5 brachydanto  
 012806 homo sapien  
 09X333 bacillus an  
 090970 rattus norv  
 096314 bacillus an  
 096314 arabidopsis  
 049681 mycobacteri  
 047103 escherichia  
 09GMD7 macaca fasc  
 050974 neisseria g  
 P73082 synecocyst  
 09UDB8 homo sapien  
 029183 sus scrofa  
 09KEX7 bacillus ha  
 018908 caenorhabdi  
 095596 plautanus c  
 09VME7 drosophila  
 086539 herpes simp  
 085938 sphingomona  
 09PC98 xyella fas  
 09RRN0 deinococcus  
 051800 alcaligenes  
 09ZTX1 bordem vul  
 09XHZ4 arabidopsis  
 09GTR0 arabidopsis  
 09SFR3 arabidopsis  
 087046 vibrio chol  
 032559 escherichia  
 09JRK8 lactococcus  
 09EYB7 escherichia  
 048423 bacterioph  
 09XJ80 bacterioph  
 09RUV1 bacterioph  
 09NUN3 deinococcus  
 09H4C2 leishmania  
 09Q0U1 homo sapien  
 09QAD8 haemophilu  
 09V489 drosophila  
 09UAE3 haemophilu  
 09UAD9 haemophilu  
 09UAD5 haemophilu  
 09UAD5 haemophilu  
 09UAD5 haemophilu  
 032897 mycobacteri  
 093442 lampetra fl  
 09YRD3 regina rana  
 09YRD2 european ca

239	6	1.7	102	13	098970	098970 oryzias lat	312	6	1.7	127	2	09EUD6	09EUD6 klebsiella
240	6	1.7	102	14	085128	085128 human parvo	313	6	1.7	127	2	09EUD6	09EUD6 klebsiella
241	6	1.7	102	14	091C7	091C7 human parvo	314	6	1.7	127	2	09EUD6	09EUD6 klebsiella
242	6	1.7	102	14	091C6	091C6 human parvo	315	6	1.7	127	2	09EUD6	09EUD6 klebsiella
243	6	1.7	102	14	091C5	091C5 human parvo	316	6	1.7	127	2	09EUD6	09EUD6 klebsiella
244	6	1.7	102	14	091C4	091C4 human parvo	317	6	1.7	127	2	09EUD6	09EUD6 klebsiella
245	6	1.7	102	14	091C3	091C3 human parvo	318	6	1.7	127	2	09EUD6	09EUD6 klebsiella
246	6	1.7	102	14	091C2	091C2 human parvo	319	6	1.7	127	2	09EUD6	09EUD6 klebsiella
247	6	1.7	102	14	091C1	091C1 human parvo	320	6	1.7	127	2	09EUD6	09EUD6 klebsiella
248	6	1.7	102	14	091B9	091B9 human parvo	321	6	1.7	127	2	09EUD6	09EUD6 klebsiella
249	6	1.7	104	4	09HOC6	09HOC6 homo sapien	322	6	1.7	130	4	09HOC6	09HOC6 homo sapien
250	6	1.7	107	6	09GK21	09GK21 ascaris suu	323	6	1.7	130	5	09GK21	09GK21 ascaris suu
251	6	1.7	107	6	09GK21	09GK21 escherichia	324	6	1.7	131	2	09HND2	09HND2 pseudomonas
252	6	1.7	107	7	09XRM5	09XRM5 phasianus c	325	6	1.7	131	2	09HND2	09HND2 pseudomonas
253	6	1.7	108	1	09Y9N3	09Y9N3 aeropyrum p	326	6	1.7	132	3	09V2D4	09V2D4 schizosacch
254	6	1.7	109	11	P70407	P70407 mus musculu	327	6	1.7	133	3	09V2D4	09V2D4 schizosacch
255	6	1.7	110	1	058605	058605 pyrococcus	328	6	1.7	133	2	09HND2	09HND2 pseudomonas
256	6	1.7	111	2	09HRS5	09HRS5 pseudomonas	329	6	1.7	133	2	09HND2	09HND2 pseudomonas
257	6	1.7	112	5	09V7M4	09V7M4 drosophila	330	6	1.7	133	2	09HND2	09HND2 pseudomonas
258	6	1.7	113	10	09LIX3	09LIX3 oryza sativ	331	6	1.7	133	2	09HND2	09HND2 pseudomonas
259	6	1.7	113	14	067411	067411 influenza c	332	6	1.7	134	2	09X7D3	09X7D3 rattus norv
260	6	1.7	114	1	057752	057752 pyrococcus	333	6	1.7	134	4	016378	016378 homo sapien
261	6	1.7	115	1	P95435	P95435 pseudomonas	334	6	1.7	134	14	085178	085178 human parvo
262	6	1.7	115	14	09Y9Q8	09Y9Q8 human eryth	335	6	1.7	135	2	050456	050456 mycobacteri
263	6	1.7	115	14	09Q2N1	09Q2N1 pseudomonas	336	6	1.7	135	2	007741	007741 mycobacteri
264	6	1.7	116	2	0916H1	0916H1 pseudomonas	337	6	1.7	136	5	018737	018737 ceanorhabdi
265	6	1.7	117	2	09KWH8	09KWH8 streptomyce	338	6	1.7	136	5	018737	018737 ceanorhabdi
266	6	1.7	118	4	09U166	09U166 homo sapien	339	6	1.7	137	4	014620	014620 homo sapien
267	6	1.7	118	5	09N9K7	09N9K7 leishmania	340	6	1.7	137	4	014620	014620 homo sapien
268	6	1.7	119	4	09N9K7	09N9K7 homo sapien	341	6	1.7	139	2	054289	054289 salmoneilla
269	6	1.7	120	14	041998	041998 viral hemor	342	6	1.7	139	2	054289	054289 salmoneilla
270	6	1.7	120	14	041999	041999 viral hemor	343	6	1.7	140	11	09R1S9	09R1S9 rattus norv
271	6	1.7	120	14	042000	042000 viral hemor	344	6	1.7	141	10	09M3V3	09M3V3 hordeum vul
272	6	1.7	120	14	042001	042001 viral hemor	345	6	1.7	141	10	09M3V3	09M3V3 hordeum vul
273	6	1.7	120	14	042002	042002 viral hemor	346	6	1.7	142	10	09M3V5	09M3V5 hordeum vul
274	6	1.7	120	14	042003	042003 viral hemor	347	6	1.7	142	10	09M3V5	09M3V5 hordeum vul
275	6	1.7	120	14	042005	042005 viral hemor	348	6	1.7	142	10	09M3V4	09M3V4 hordeum vul
276	6	1.7	120	14	042009	042009 viral hemor	349	6	1.7	142	10	09M3V4	09M3V4 hordeum vul
277	6	1.7	120	14	042010	042010 viral hemor	350	6	1.7	143	5	021524	021524 ceanorhabdi
278	6	1.7	120	14	042011	042011 viral hemor	351	6	1.7	143	5	021524	021524 ceanorhabdi
279	6	1.7	120	14	042012	042012 viral hemor	352	6	1.7	143	5	021524	021524 ceanorhabdi
280	6	1.7	120	14	042013	042013 viral hemor	353	6	1.7	143	5	021524	021524 ceanorhabdi
281	6	1.7	120	14	042014	042014 viral hemor	354	6	1.7	143	5	021524	021524 ceanorhabdi
282	6	1.7	120	14	042015	042015 viral hemor	355	6	1.7	143	5	021524	021524 ceanorhabdi
283	6	1.7	120	14	042016	042016 viral hemor	356	6	1.7	143	5	021524	021524 ceanorhabdi
284	6	1.7	120	14	042017	042017 viral hemor	357	6	1.7	143	5	021524	021524 ceanorhabdi
285	6	1.7	120	14	042019	042019 viral hemor	358	6	1.7	143	5	021524	021524 ceanorhabdi
286	6	1.7	120	14	042020	042020 viral hemor	359	6	1.7	143	5	021524	021524 ceanorhabdi
287	6	1.7	120	14	042021	042021 viral hemor	360	6	1.7	143	5	021524	021524 ceanorhabdi
288	6	1.7	120	14	042022	042022 viral hemor	361	6	1.7	143	5	021524	021524 ceanorhabdi
289	6	1.7	120	14	042023	042023 viral hemor	362	6	1.7	143	5	021524	021524 ceanorhabdi
290	6	1.7	120	14	042024	042024 viral hemor	363	6	1.7	143	5	021524	021524 ceanorhabdi
291	6	1.7	120	14	042025	042025 viral hemor	364	6	1.7	143	5	021524	021524 ceanorhabdi
292	6	1.7	120	14	087065	087065 pseudorhab	365	6	1.7	148	10	09M4E1	09M4E1 avena sativ
293	6	1.7	122	2	09RCS0	09RCS0 bacillus ha	366	6	1.7	148	10	09M4E1	09M4E1 avena sativ
294	6	1.7	123	2	09KVI1	09KVI1 vibrio chol	367	6	1.7	148	10	09M4E0	09M4E0 triticum mo
295	6	1.7	123	2	051512	051512 pseudomonas	368	6	1.7	148	10	09M4D9	09M4D9 triticum mo
296	6	1.7	125	14	09Q2S2	09Q2S2 influenza c	369	6	1.7	149	8	09G827	09G827 gymnotus pa
297	6	1.7	125	14	09Q249	09Q249 influenza c	370	6	1.7	149	8	09G827	09G827 gymnotus pa
298	6	1.7	125	14	09Q246	09Q246 influenza c	371	6	1.7	149	8	09G827	09G827 gymnotus pa
299	6	1.7	125	14	09Q243	09Q243 influenza c	372	6	1.7	149	8	09G827	09G827 gymnotus pa
300	6	1.7	125	14	09Q234	09Q234 influenza c	373	6	1.7	149	8	09G827	09G827 gymnotus pa
301	6	1.7	125	14	09Q220	09Q220 influenza c	374	6	1.7	150	2	09J231	09J231 neisseria m
302	6	1.7	126	13	028397	028397 archaoglob	375	6	1.7	150	2	09J231	09J231 neisseria m
303	6	1.7	126	13	091442	091442 squatus aca	376	6	1.7	150	2	09J231	09J231 neisseria m
304	6	1.7	127	2	09E278	09E278 klebsiella	377	6	1.7	150	5	018574	018574 buena sp.
305	6	1.7	127	2	09E277	09E277 klebsiella	378	6	1.7	151	4	09H4C4	09H4C4 homo sapien
306	6	1.7	127	2	09E276	09E276 klebsiella	379	6	1.7	151	4	09H4C4	09H4C4 homo sapien
307	6	1.7	127	2	09E275	09E275 klebsiella	380	6	1.7	151	4	09H4C4	09H4C4 homo sapien
308	6	1.7	127	2	09E274	09E274 klebsiella	381	6	1.7	151	4	09H4C4	09H4C4 homo sapien
309	6	1.7	127	2	09E273	09E273 klebsiella	382	6	1.7	151	4	09H4C4	09H4C4 homo sapien
310	6	1.7	127	2	09E272	09E272 klebsiella	383	6	1.7	151	4	09H4C4	09H4C4 homo sapien
311	6	1.7	127	2	09E270	09E270 klebsiella	384	6	1.7	151	4	09H4C4	09H4C4 homo sapien

385	6	1.7	155	2	P72677	P72677 synecocyst	458	6	1.7	186	14	085132	085132 human parvo
386	6	1.7	156	1	059154	059154 pyrococcus	459	6	1.7	186	14	085163	085163 human parvo
387	6	1.7	158	2	P94509	P94509 bacillus su	460	6	1.7	187	14	085163	085163 human parvo
388	6	1.7	158	2	09PR82	09PR82 ureaplasma v	461	6	1.7	187	14	09WRT6	09WRT6 streptomyc
389	6	1.7	159	10	081371	081371 phaseolus v	462	6	1.7	189	5	09WRT6	09WRT6 streptomyc
390	6	1.7	159	10	09LIXO	09LIXO phaseolus v	463	6	1.7	189	5	09WRT6	09WRT6 streptomyc
391	6	1.7	159	10	09LIXO	09LIXO phaseolus v	464	6	1.7	190	10	004414	004414 phaseolus v
392	6	1.7	160	8	09XMA4	09XMA4 phaseolus v	465	6	1.7	190	10	004405	004405 phaseolus v
393	6	1.7	160	8	09XMA4	09XMA4 phaseolus v	466	6	1.7	190	10	004405	004405 phaseolus v
394	6	1.7	161	1	09YBU0	09YBU0 phaseolus v	467	6	1.7	190	10	004405	004405 phaseolus v
395	6	1.7	161	2	085683	085683 streptomyc	468	6	1.7	190	10	004405	004405 phaseolus v
396	6	1.7	161	2	09BVA3	09BVA3 streptomyc	469	6	1.7	191	2	09WRT6	09WRT6 streptomyc
397	6	1.7	161	2	09BVA3	09BVA3 streptomyc	470	6	1.7	191	2	09WRT6	09WRT6 streptomyc
398	6	1.7	161	13	09IB26	09IB26 trachurus t	471	6	1.7	192	2	09WRT6	09WRT6 streptomyc
399	6	1.7	162	8	09BGR7	09BGR7 homo sapien	472	6	1.7	192	2	09WRT6	09WRT6 streptomyc
400	6	1.7	162	8	09BGR7	09BGR7 homo sapien	473	6	1.7	193	2	09WRT6	09WRT6 streptomyc
401	6	1.7	163	4	09NRJ2	09NRJ2 homo sapien	474	6	1.7	194	2	09WRT6	09WRT6 streptomyc
402	6	1.7	163	4	09NRJ2	09NRJ2 homo sapien	475	6	1.7	194	2	09WRT6	09WRT6 streptomyc
403	6	1.7	163	10	09FXC5	09FXC5 streptomyc	476	6	1.7	195	2	09WRT6	09WRT6 streptomyc
404	6	1.7	164	2	054209	054209 streptomyc	477	6	1.7	195	2	09WRT6	09WRT6 streptomyc
405	6	1.7	164	2	09RDP6	09RDP6 streptomyc	478	6	1.7	195	2	09WRT6	09WRT6 streptomyc
406	6	1.7	166	2	031769	031769 bacillus su	479	6	1.7	196	2	032394	032394 klebsiella
407	6	1.7	166	5	09VNR4	09VNR4 drosophila	480	6	1.7	196	2	032394	032394 klebsiella
408	6	1.7	168	1	058698	058698 mechanobact	481	6	1.7	196	11	054918	054918 mus musculu
409	6	1.7	169	1	027384	027384 streptomyc	482	6	1.7	197	2	032146	032146 bacillus su
410	6	1.7	169	1	09XAU4	09XAU4 streptomyc	483	6	1.7	197	2	032146	032146 bacillus su
411	6	1.7	169	10	09SYD5	09SYD5 arabisopsis	484	6	1.7	197	10	09SDNO	09SDNO arabisopsis
412	6	1.7	170	8	09TRK5	09TRK5 echinococcu	485	6	1.7	197	10	09SDNO	09SDNO arabisopsis
413	6	1.7	170	8	09TRK5	09TRK5 echinococcu	486	6	1.7	197	10	09SDNO	09SDNO arabisopsis
414	6	1.7	171	2	09WZP1	09WZP1 thermotoga	487	6	1.7	198	14	043521	043521 equine herp
415	6	1.7	171	5	09NEB8	09NEB8 leishmania	488	6	1.7	198	14	043521	043521 equine herp
416	6	1.7	173	2	09ISJ2	09ISJ2 chlamydia p	489	6	1.7	199	2	09L516	09L516 pseudomonas
417	6	1.7	174	2	053746	053746 mycobacteri	490	6	1.7	200	2	09WRT6	09WRT6 streptomyc
418	6	1.7	174	10	P93571	P93571 solanum tub	491	6	1.7	203	5	019759	019759 pseudomonas
419	6	1.7	175	2	087797	087797 pseudomonas	492	6	1.7	204	2	09K182	09K182 human parvo
420	6	1.7	175	11	09WV79	09WV79 rattus norv	493	6	1.7	204	2	09K182	09K182 human parvo
421	6	1.7	176	2	P72779	P72779 synecocyst	494	6	1.7	205	2	09JUB6	09JUB6 pseudomonas
422	6	1.7	176	2	09X2H7	09X2H7 thermotoga	495	6	1.7	205	2	09JUB6	09JUB6 pseudomonas
423	6	1.7	176	2	09X2H7	09X2H7 thermotoga	496	6	1.7	205	2	09JUB6	09JUB6 pseudomonas
424	6	1.7	176	14	085182	085182 human parvo	497	6	1.7	206	1	09YDE6	09YDE6 streptomyc
425	6	1.7	177	1	09YCX1	09YCX1 aetopyrum p	498	6	1.7	206	2	005747	005747 drosophila
426	6	1.7	177	2	09K155	09K155 streptomyc	499	6	1.7	206	2	005747	005747 drosophila
427	6	1.7	177	5	020153	020153 caenorhabdi	500	6	1.7	207	2	032241	032241 bacillus su
428	6	1.7	178	2	09R095	09R095 delinococcus	501	6	1.7	207	2	032241	032241 bacillus su
429	6	1.7	179	2	09WNT3	09WNT3 streptomyc	502	6	1.7	212	2	025248	025248 heliobacte
430	6	1.7	179	2	09WNT3	09WNT3 streptomyc	503	6	1.7	212	2	025248	025248 heliobacte
431	6	1.7	179	10	09XG03	09XG03 oryza sativ	504	6	1.7	212	2	09XG03	09XG03 oryza sativ
432	6	1.7	179	14	085174	085174 human parvo	505	6	1.7	212	10	09XG03	09XG03 oryza sativ
433	6	1.7	180	2	09XG03	09XG03 oryza sativ	506	6	1.7	213	1	09XG03	09XG03 oryza sativ
434	6	1.7	180	3	074882	074882 schizosacch	507	6	1.7	213	2	09XG03	09XG03 oryza sativ
435	6	1.7	180	10	09L6C7	09L6C7 oryza sativ	508	6	1.7	213	2	09XG03	09XG03 oryza sativ
436	6	1.7	180	10	09L6C7	09L6C7 oryza sativ	509	6	1.7	213	2	09XG03	09XG03 oryza sativ
437	6	1.7	182	5	09Y1Y6	09Y1Y6 ephydaria f	510	6	1.7	214	10	039763	039763 gossypium b
438	6	1.7	182	8	09Y1Y6	09Y1Y6 ephydaria f	511	6	1.7	214	10	039763	039763 gossypium b
439	6	1.7	182	10	09Z095	09Z095 cyanidium c	512	6	1.7	214	10	039763	039763 gossypium b
440	6	1.7	182	13	09Z095	09Z095 cyanidium c	513	6	1.7	214	10	039763	039763 gossypium b
441	6	1.7	182	13	09Z095	09Z095 cyanidium c	514	6	1.7	214	10	039763	039763 gossypium b
442	6	1.7	182	13	09Z095	09Z095 cyanidium c	515	6	1.7	215	14	09XG03	09XG03 oryza sativ
443	6	1.7	182	13	09Z095	09Z095 cyanidium c	516	6	1.7	215	14	09XG03	09XG03 oryza sativ
444	6	1.7	182	13	09Z095	09Z095 cyanidium c	517	6	1.7	215	13	073722	073722 orochromis
445	6	1.7	183	14	085147	085147 human parvo	518	6	1.7	215	13	073722	073722 orochromis
446	6	1.7	183	14	085147	085147 human parvo	519	6	1.7	215	13	073722	073722 orochromis
447	6	1.7	183	2	056960	056960 yersinia ps	520	6	1.7	216	1	09H056	09H056 lates calca
448	6	1.7	183	2	09XAV3	09XAV3 pseudomonas	521	6	1.7	216	1	09H056	09H056 lates calca
449	6	1.7	183	5	09VRI3	09VRI3 pseudomonas	522	6	1.7	216	10	09XMA4	09XMA4 phaseolus v
450	6	1.7	184	5	09VRI3	09VRI3 pseudomonas	523	6	1.7	217	10	09XMA4	09XMA4 phaseolus v
451	6	1.7	184	5	09VRI3	09VRI3 pseudomonas	524	6	1.7	218	2	007387	007387 lactococcus
452	6	1.7	184	5	09VRI3	09VRI3 pseudomonas	525	6	1.7	218	2	007387	007387 lactococcus
453	6	1.7	184	5	09VRI3	09VRI3 pseudomonas	526	6	1.7	219	5	09WRT6	09WRT6 streptomyc
454	6	1.7	185	2	09RNR8	09RNR8 drosophila	527	6	1.7	219	5	09WRT6	09WRT6 streptomyc
455	6	1.7	185	2	09RNR8	09RNR8 drosophila	528	6	1.7	219	5	09WRT6	09WRT6 streptomyc
456	6	1.7	185	5	022535	022535 caenorhabdi	529	6	1.7	220	8	09MD15	09MD15 sceneemus
457	6	1.7	186	2	09L267	09L267 streptomyc	530	6	1.7	220	8	09MD15	09MD15 sceneemus



677	6	1.7	258	5	09GR09	09gr09 leishmania	750	6	1.7	280	8	09TNU3	09tnj3 haplomitriu
678	6	1.7	258	14	08GR75	08GR75 cowpea molt	751	6	1.7	281	1	030017	030017 archaeoglob
679	6	1.7	258	2	09Z460	09Z460 paracoccus	752	6	1.7	281	1	09YCZ8	09ycz8 aeropyrum p
680	6	1.7	258	2	059694	059694 pseudomonas	753	6	1.7	281	2	09K4E3	09k4e3 streptomyc
681	6	1.7	258	2	059694	059694 pseudomonas	754	6	1.7	281	2	09K4E3	09k4e3 streptomyc
682	6	1.7	258	2	09K5E6	09K5E6 klebsiella	755	6	1.7	282	5	09VKH5	09vkh5 streptomyc
683	6	1.7	258	2	09K5E6	09K5E6 corynebacte	756	6	1.7	282	5	09VKH5	09vkh5 streptomyc
684	6	1.7	258	2	09J094	09J094 neisseria m	757	6	1.7	282	13	09DPL1	09ddl1 gallus gall
685	6	1.7	258	2	051348	051348 pseudomonas	757	6	1.7	283	1	028913	028913 archaeoglob
686	6	1.7	258	2	046620	046620 erwania amy	758	6	1.7	283	2	09RA04	09ra04 rhodococcus
687	6	1.7	258	2	09K8T6	09K8T6 streptomyc	759	6	1.7	283	2	09PMN1	09pmn1 campylobact
688	6	1.7	258	5	09M1T5	09M1T5 streptomyc	760	6	1.7	283	2	09K7J6	09k7j6 bacillus ha
689	6	1.7	258	5	09U272	09U272 caenorhabdi	761	6	1.7	283	10	048807	048807 arabisdopsi
690	6	1.7	258	2	09KDC0	09KDC0 streptomyc	762	6	1.7	284	2	09KX18	09kx18 streptomyc
691	6	1.7	258	2	09P26	09P26 campylobact	763	6	1.7	285	2	09KX18	09kx18 streptomyc
692	6	1.7	258	2	09K7R7	09K7R7 bacillus ha	764	6	1.7	285	3	044599	044599 saccharomyc
693	6	1.7	258	2	09H772	09H772 pseudomonas	765	6	1.7	286	2	007469	007469 rhodopseudo
694	6	1.7	258	2	09S105	09S105 homo sapien	766	6	1.7	286	2	053409	053409 mycoplasma
695	6	1.7	258	2	070004	070004 plasamid nr7	767	6	1.7	286	2	09KQ26	09kq26 vibrio chol
696	6	1.7	258	2	09ZHE5	09ZHE5 salmonella	768	6	1.7	286	5	017670	017670 caenorhabdi
697	6	1.7	258	2	09K1U8	09K1U8 vibrio chol	769	6	1.7	286	5	076137	076137 drosophila
698	6	1.7	258	2	09K1R2	09K1R2 klebsiella	770	6	1.7	286	5	09UA14	09ua14 drosophila
699	6	1.7	258	2	09K3E0	09K3E0 salmonella	771	6	1.7	286	10	038910	038910 arabisdopsi
700	6	1.7	258	2	09F7C0	09F7C0 serratia ma	772	6	1.7	287	1	09YC84	09yc84 aeropyrum p
701	6	1.7	258	2	084819	084819 xenopus lae	773	6	1.7	287	10	09Y57	09y57 pyrococcus
702	6	1.7	258	2	084819	084819 xenopus lae	774	6	1.7	287	10	09ERN2	09ern2 oryza sativ
703	6	1.7	258	2	09LDJ0	09LDJ0 oryza sativ	775	6	1.7	287	13	091002	091002 gallus gall
704	6	1.7	258	14	069053	069053 human herpe	776	6	1.7	287	14	090758	090758 fowlpox vir
705	6	1.7	258	8	09MFS4	09MFS4 magnaia gr	777	6	1.7	288	2	09X246	09x246 escherichia
706	6	1.7	258	8	09MFS4	09MFS4 magnaia gr	778	6	1.7	288	2	09X246	09x246 thermotoga
707	6	1.7	258	10	081213	081213 zea mays (m	779	6	1.7	289	10	041216	041216 trichosanth
708	6	1.7	258	10	09FEV4	09FEV4 olea europ	780	6	1.7	289	10	039563	039563 closterium
709	6	1.7	258	2	09HXP7	09HXP7 pseudomonas	781	6	1.7	290	8	09TN12	09tn12 physoclitre
710	6	1.7	258	2	09HXP7	09HXP7 pseudomonas	782	6	1.7	291	2	083025	083025 enterobacte
711	6	1.7	258	4	09VNY1	09VNY1 homo sapien	783	6	1.7	291	2	084829	084829 chlamydia t
712	6	1.7	258	4	09VNY1	09VNY1 homo sapien	784	6	1.7	291	2	09PL98	09pl98 chlamydia m
713	6	1.7	258	5	09G1E5	09G1E5 pyrococcus	785	6	1.7	291	2	09PL98	09pl98 chlamydia m
714	6	1.7	258	5	09G1E5	09G1E5 pyrococcus	786	6	1.7	291	2	09PL98	09pl98 chlamydia m
715	6	1.7	258	5	09V4A9	09V4A9 pseudomonas	787	6	1.7	291	2	09PA92	09pa92 enterobacte
716	6	1.7	258	5	09V4A9	09V4A9 pseudomonas	788	6	1.7	291	2	09PA92	09pa92 enterobacte
717	6	1.7	258	1	027468	027468 methanobact	789	6	1.7	291	2	09EX98	09ex98 ochrobactru
718	6	1.7	258	1	059447	059447 pyrococcus	790	6	1.7	291	8	09ET55	09et55 parides hab
719	6	1.7	258	13	09H873	09H873 pyrococcus	791	6	1.7	291	8	09ET55	09et55 parides hab
720	6	1.7	258	13	09H873	09H873 pyrococcus	792	6	1.7	291	8	09ET55	09et55 parides hab
721	6	1.7	258	2	09HXP7	09HXP7 pseudomonas	793	6	1.7	291	10	09FRB2	09frb2 oryza sativ
722	6	1.7	258	2	09HXP7	09HXP7 pseudomonas	794	6	1.7	292	1	09HJ17	09hj17 thermoplasma
723	6	1.7	258	4	09HBL3	09HBL3 schizosacch	795	6	1.7	292	5	008088	008088 tritrichomo
724	6	1.7	258	4	09HBL3	09HBL3 schizosacch	796	6	1.7	292	5	029843	029843 archaeoglob
725	6	1.7	258	5	09X4T2	09X4T2 streptococc	797	6	1.7	293	2	09ZFR5	09zfr5 chlamydia p
726	6	1.7	258	5	09X4T2	09X4T2 streptococc	798	6	1.7	293	13	09YGS3	09ygs3 fuigu rubrip
727	6	1.7	258	8	063308	063308 pleris rupa	799	6	1.7	294	2	052851	052851 bacillus pu
728	6	1.7	258	8	063308	063308 pleris rupa	800	6	1.7	294	3	013851	013851 schizosacch
729	6	1.7	258	8	063308	063308 pleris rupa	801	6	1.7	294	4	09NO60	09no60 homo sapien
730	6	1.7	258	8	063308	063308 pleris rupa	802	6	1.7	294	4	09H917	09h917 homo sapien
731	6	1.7	258	8	063308	063308 pleris rupa	803	6	1.7	295	5	020242	020242 caenorhabdi
732	6	1.7	258	8	063308	063308 pleris rupa	804	6	1.7	295	5	09NH07	09nh07 heliobact
733	6	1.7	258	10	022400	022400 arabisdopsi	805	6	1.7	295	10	09VFR6	09vfr6 arabisdopsi
734	6	1.7	258	10	022400	022400 arabisdopsi	806	6	1.7	297	2	09ZB12	09zb12 chlamydia p
735	6	1.7	258	10	022400	022400 arabisdopsi	807	6	1.7	297	2	09ZB12	09zb12 chlamydia p
736	6	1.7	258	13	09P26	09P26 campylobact	808	6	1.7	297	2	09ZB12	09zb12 chlamydia p
737	6	1.7	258	13	09P26	09P26 campylobact	809	6	1.7	299	2	09JMK1	09jmk1 neisseria m
738	6	1.7	258	13	09P26	09P26 campylobact	810	6	1.7	299	2	09JMK1	09jmk1 neisseria m
739	6	1.7	258	13	09P26	09P26 campylobact	811	6	1.7	300	2	09WTC3	09wtc3 streptomyc
740	6	1.7	258	13	09P26	09P26 campylobact	812	6	1.7	300	5	09NHO8	09nho8 agrotis ips
741	6	1.7	258	13	09P26	09P26 campylobact	813	6	1.7	301	2	09KSV5	09ksv5 arabisdopsi
742	6	1.7	258	13	09P26	09P26 campylobact	814	6	1.7	301	2	09KSV5	09ksv5 arabisdopsi
743	6	1.7	258	13	09P26	09P26 campylobact	815	6	1.7	301	2	09KSV5	09ksv5 arabisdopsi
744	6	1.7	258	13	09P26	09P26 campylobact	816	6	1.7	301	2	09KSV5	09ksv5 arabisdopsi
745	6	1.7	258	13	09P26	09P26 campylobact	817	6	1.7	301	2	09KSV5	09ksv5 arabisdopsi
746	6	1.7	258	13	09P26	09P26 campylobact	818	6	1.7	301	2	09KSV5	09ksv5 arabisdopsi
747	6	1.7	258	13	09P26	09P26 campylobact	819	6	1.7	301	2	09KSV5	09ksv5 arabisdopsi
748	6	1.7	258	13	09P26	09P26 campylobact	820	6	1.7	301	2	09KSV5	09ksv5 arabisdopsi
749	6	1.7	258	13	09P26	09P26 campylobact	821	6	1.7	301	2	09KSV5	09ksv5 arabisdopsi
	6	1.7	258	13	09P26	09P26 campylobact	822	6	1.7	301	2	09KSV5	09ksv5 arabisdopsi



823	6	1.7	303	10	041814	041814 zea mays (m	896	6	1.7	324	2	09Z4H4	09Z4H4 serratia ma
824	6	1.7	304	2	09RZJ6	09RZJ6 delinococcus	897	6	1.7	324	2	09JXL0	09JXL0 neisseria m
825	6	1.7	304	10	09SYE1	09SYE1 aradiopsis	898	6	1.7	324	8	09GFP0	09GFP0 illicium pa
826	6	1.7	305	1	09HQ06	09HQ06 halobacteri	899	6	1.7	324	8	09GFE6	09GFE6 lithodendr
827	6	1.7	305	2	067486	067486 aquilex aeo	900	6	1.7	324	8	09GFE2	09GFE2 trochodendr
828	6	1.7	305	2	007166	007166 mycobacteri	901	6	1.7	324	10	049662	049662 aradiopsis
829	6	1.7	305	2	033477	033477 pseudomonas	902	6	1.7	324	13	09U498	09U498 myxine gut
830	6	1.7	305	2	0914V1	0914V1 pseudomonas	903	6	1.7	324	14	037399	037399 chlorella v
831	6	1.7	305	2	052740	052740 corynebacte	904	6	1.7	325	8	09GRC2	09GRC2 ceratophyll
832	6	1.7	305	2	052740	052740 corynebacte	905	6	1.7	325	8	09GRC2	09GRC2 ceratophyll
833	6	1.7	306	2	050798	050798 borrelia bu	906	6	1.7	325	8	09GRC0	09GRC0 iactoris te
834	6	1.7	306	2	050798	050798 borrelia bu	907	6	1.7	325	8	09GRC0	09GRC0 iactoris te
835	6	1.7	307	2	09RUT5	09RUT5 neisseria m	908	6	1.7	326	2	09GPA6	09GPA6 uncultured
836	6	1.7	307	2	09RUT5	09RUT5 neisseria m	909	6	1.7	326	2	09GPA6	09GPA6 uncultured
837	6	1.7	307	2	09RUT5	09RUT5 neisseria m	910	6	1.7	326	2	09GPA6	09GPA6 uncultured
838	6	1.7	307	14	09Q3H0	09Q3H0 west nile v	911	6	1.7	326	2	051945	051945 neisseria g
839	6	1.7	307	14	09Q3H0	09Q3H0 west nile v	912	6	1.7	326	2	051945	051945 neisseria g
840	6	1.7	307	14	09Q3H0	09Q3H0 west nile v	913	6	1.7	326	2	051948	051948 neisseria g
841	6	1.7	308	2	09X9J3	09X9J3 vibrio para	914	6	1.7	326	2	09Z4K8	09Z4K8 neisseria g
842	6	1.7	308	4	09UPU3	09UPU3 homo sapien	915	6	1.7	326	2	09Z4K1	09Z4K1 neisseria g
843	6	1.7	308	5	076762	076762 anopheles g	916	6	1.7	326	2	051944	051944 neisseria g
844	6	1.7	308	10	040668	040668 oryza sativ	917	6	1.7	326	2	09RKP6	09RKP6 streptomyce
845	6	1.7	309	2	054023	054023 paracoccus	918	6	1.7	326	2	09RKP6	09RKP6 streptomyce
846	6	1.7	310	5	023059	023059 caenorhabdi	919	6	1.7	326	2	09RKP6	09RKP6 streptomyce
847	6	1.7	311	2	082874	082874 streptococ	920	6	1.7	326	2	09RKP6	09RKP6 streptomyce
848	6	1.7	311	10	09LON9	09LON9 aradiopsis	921	6	1.7	326	2	09RKP6	09RKP6 streptomyce
849	6	1.7	312	5	061848	061848 caenorhabdi	922	6	1.7	326	2	09RKP6	09RKP6 streptomyce
850	6	1.7	312	5	09V6E9	09V6E9 dirosophila	923	6	1.7	326	2	09RKP6	09RKP6 streptomyce
851	6	1.7	312	8	09G3U3	09G3U3 pneumocysti	924	6	1.7	326	2	09RKP6	09RKP6 streptomyce
852	6	1.7	312	8	09G3U1	09G3U1 mouse cytom	925	6	1.7	326	2	09RKP6	09RKP6 streptomyce
853	6	1.7	312	14	069217	069217 agrobacteri	926	6	1.7	326	2	09RKP6	09RKP6 streptomyce
854	6	1.7	313	3	09R7U0	09R7U0 agrobacteri	927	6	1.7	326	2	09RKP6	09RKP6 streptomyce
855	6	1.7	313	3	09R7U0	09R7U0 agrobacteri	928	6	1.7	326	2	09RKP6	09RKP6 streptomyce
856	6	1.7	313	10	09MOM0	09MOM0 aradiopsis	929	6	1.7	327	1	09R070	09R070 halobacteri
857	6	1.7	314	4	09NKH1	09NKH1 homo sapien	930	6	1.7	327	1	09R070	09R070 halobacteri
858	6	1.7	314	8	09MM08	09MM08 anthoceros	931	6	1.7	328	2	086149	086149 campylobact
859	6	1.7	314	10	043561	043561 medicago tr	932	6	1.7	328	2	09R1Y4	09R1Y4 streptomyce
860	6	1.7	315	2	09JWC1	09JWC1 neisseria m	933	6	1.7	328	2	09R1Y4	09R1Y4 streptomyce
861	6	1.7	315	4	09NHS1	09NHS1 homo sapien	934	6	1.7	328	2	09R1Y4	09R1Y4 streptomyce
862	6	1.7	316	11	09QX02	09QX02 mus musculu	935	6	1.7	328	2	09R1Y4	09R1Y4 streptomyce
863	6	1.7	316	11	09QX02	09QX02 mus musculu	936	6	1.7	328	2	09R1Y4	09R1Y4 streptomyce
864	6	1.7	316	14	09XES9	09XES9 ryegrass mo	937	6	1.7	328	2	09R1Y4	09R1Y4 streptomyce
865	6	1.7	316	14	09XES9	09XES9 ryegrass mo	938	6	1.7	328	2	09R1Y4	09R1Y4 streptomyce
866	6	1.7	317	2	048981	048981 mycoplasma	939	6	1.7	329	10	043562	043562 campylobact
867	6	1.7	317	2	048981	048981 mycoplasma	940	6	1.7	329	10	043562	043562 campylobact
868	6	1.7	317	3	09UWD9	09UWD9 streptomyce	941	6	1.7	330	2	09K5D5	09K5D5 campylobact
869	6	1.7	317	3	09UWD9	09UWD9 streptomyce	942	6	1.7	330	2	09K5D5	09K5D5 campylobact
870	6	1.7	317	14	086175	086175 rotavirus s	943	6	1.7	330	2	09K5D5	09K5D5 campylobact
871	6	1.7	318	2	055862	055862 streptocyst	944	6	1.7	331	2	09K5D5	09K5D5 campylobact
872	6	1.7	318	2	09XAK3	09XAK3 pseudomonas	945	6	1.7	331	2	09K5D5	09K5D5 campylobact
873	6	1.7	318	2	09XAK3	09XAK3 pseudomonas	946	6	1.7	331	2	09K5D5	09K5D5 campylobact
874	6	1.7	319	2	09S265	09S265 mycobacteri	947	6	1.7	331	2	09K5D5	09K5D5 campylobact
875	6	1.7	319	4	09UFR3	09UFR3 homo sapien	948	6	1.7	331	2	09K5D5	09K5D5 campylobact
876	6	1.7	319	8	09RUT9	09RUT9 sus scrofa	949	6	1.7	332	4	09RKF7	09RKF7 coleochaete
877	6	1.7	319	8	09RUT9	09RUT9 sus scrofa	950	6	1.7	332	4	09RKF7	09RKF7 coleochaete
878	6	1.7	319	9	09KCM2	09KCM2 streptococ	951	6	1.7	333	10	09S238	09S238 ratiuss novr
879	6	1.7	319	10	09KCM2	09KCM2 streptococ	952	6	1.7	333	10	09S238	09S238 ratiuss novr
880	6	1.7	320	2	000337	000337 salinonella	953	6	1.7	333	10	09S238	09S238 ratiuss novr
881	6	1.7	320	4	09UHX6	09UHX6 homo sapien	954	6	1.7	334	5	018353	018353 caenorhabdi
882	6	1.7	320	5	076487	076487 dirosophila	955	6	1.7	334	5	018353	018353 caenorhabdi
883	6	1.7	320	10	042994	042994 oryza sativ	956	6	1.7	334	10	09M422	09M422 spinctia ol
884	6	1.7	321	8	09GFF8	09GFF8 dioscorea b	957	6	1.7	334	10	09M422	09M422 spinctia ol
885	6	1.7	322	1	027145	027145 methanobact	958	6	1.7	334	10	09M422	09M422 spinctia ol
886	6	1.7	322	2	09WXT4	09WXT4 thermotoga	959	6	1.7	335	2	09RNC8	09RNC8 halobacteri
887	6	1.7	322	10	09SNO9	09SNO9 aradiopsis	960	6	1.7	335	2	09RNC8	09RNC8 halobacteri
888	6	1.7	322	14	09Q8S6	09Q8S6 myxoma viru	961	6	1.7	335	2	09RNC8	09RNC8 halobacteri
889	6	1.7	323	2	09Z4O8	09Z4O8 streptomyce	962	6	1.7	335	2	09RNC8	09RNC8 halobacteri
890	6	1.7	323	5	049240	049240 homo sapien	963	6	1.7	335	2	09RNC8	09RNC8 halobacteri
891	6	1.7	323	5	049240	049240 homo sapien	964	6	1.7	336	2	09RNC8	09RNC8 halobacteri
892	6	1.7	323	8	09TUS1	09TUS1 cyanidium c	965	6	1.7	336	2	09RNC8	09RNC8 halobacteri
893	6	1.7	323	8	09GFC8	09GFC8 asarum cana	966	6	1.7	336	4	09H0N9	09H0N9 homo sapien
894	6	1.7	323	8	09GFC8	09GFC8 asarum cana	967	6	1.7	336	4	09H0N9	09H0N9 homo sapien
895	6	1.7	324	1	09HP66	09HP66 halobacteri	968	6	1.7	337	2	09R294	09R294 delinococcus

```

569 6 1.7 337 5 045734 045734 caenorhabdi
570 6 1.7 337 5 09Y156 09Y156 drosophila
571 6 1.7 337 14 055793 055793 batu cave v
572 6 1.7 337 14 055818 055818 phnom-penh
573 6 1.7 338 1 09YB01 09YB01 aetopyrum p
574 6 1.7 338 2 09RT35 09RT35 deliococcus
575 6 1.7 338 10 09FN74 09FN74 arabidopsis
576 6 1.7 339 3 09UT08 09UT08 schizosacch
577 6 1.7 339 5 026636 026636 sarcophaga
578 6 1.7 339 14 056866 056866 gallid herp
579 6 1.7 340 2 09X452 09X452 pseudomonas
580 6 1.7 340 2 09RT47 09RT47 streptomyc
581 6 1.7 340 10 09SAV3 09SAV3 oryza sativ
582 6 1.7 341 2 050655 050655 selenomonas
583 6 1.7 341 2 082857 082857 acetobacter
584 6 1.7 341 4 09UD87 09UD87 homo sapien
585 6 1.7 341 5 094703 094703 physarum po
586 6 1.7 341 8 031977 031977 cordyleclad
587 6 1.7 341 10 09LYM5 09LYM5 arabidopsis
588 6 1.7 341 10 09RJPS 09RJPS arabidopsis
589 6 1.7 342 5 08XAX0 08XAX0 aetopyrum p
590 6 1.7 342 5 018071 018071 caenorhabdi
591 6 1.7 342 5 09GT65 09GT65 drosophila
592 6 1.7 343 1 028418 028418 archaeoglob
593 6 1.7 343 5 09KD74 09KD74 bacillus ha
594 6 1.7 343 5 0917J4 0917J4 drosophila
595 6 1.7 343 8 09G620 09G620 ceratophora
596 6 1.7 343 14 055839 055839 zika virus
597 6 1.7 344 2 0925C1 0925C1 streptomyc
598 6 1.7 344 2 09RD24 09RD24 streptomyc
599 6 1.7 344 5 061754 061754 caenorhabdi
1000 6 1.7 344 10 09SQH2 09sqh2 brassica na

```

## ALIGNMENTS

```

RESULT 1
Q9H2P8 PRELIMINARY: PRT: 350 AA.
AC Q9H2P8:
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE HT028.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HYPOPHALAMUS;
RA Yang Y., Xu X., Gao G., Xiao H., Chen Z., Han Z.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF248966; ANG44564.1; -;
SQ SEQUENCE 350 AA; 39036 MW; 840A989ACC4E8D8 CRC64;

```

```

Query Match 81.1%; Score 284; DB 4; Length 350;
Best Local Similarity 100.0%; Pred. No. 4.8e-275;
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MAVPVLLALVAGVGNFESTLKSPGVSFRNGNMPDGERIPDVALSMGVSVEDLSW 60
DB 1 MAVPVLLALVAGVGNFESTLKSPGVSFRNGNMPDGERIPDVALSMGVSVEDLSW 60
QY 61 PGLAVGMLFRRPRATVWVKGKVKLALPPGVSISYPLENAVPSLDSVANSIHSLFSEE 120
DB 61 PGLAVGMLFRRPRATVWVKGKVKLALPPGVSISYPLENAVPSLDSVANSIHSLFSEE 120
QY 121 TPVVLQALASEERVMWKGANSVEDLSYTLQOLNRRIFOENSVLSLPLNSLNNEVD 160
DB 121 TPVVLQALASEERVMWKGANSVEDLSYTLQOLNRRIFOENSVLSLPLNSLNNEVD 160

```

```

QY 181 LFLSELQVLDHDISSLSRHKHLANDHSPDLYSLAGLDEIGKRYGDSSEFRDASKITL 240
DB 181 LFLSELQVLDHDISSLSRHKHLANDHSPDLYSLAGLDEIGKRYGDSSEFRDASKITL 240
QY 241 VDALKRFADDMYSILYGNMVELVTKSEFDSILRKTRITILEAK 284
DB 241 VDALKRFADDMYSILYGNMVELVTKSEFDSILRKTRITILEAK 284

```

```

RESULT 2
Q9UC89 PRELIMINARY: PRT: 330 AA.
AC Q9UC89:
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE HYPOTHEICAL 37.0 KDA PROTEIN (FRAGMENT).
GN DKF256400582.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Poustka A., Klein M., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL049929; CAB43210.1; -;
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 330 AA; 37004 MW; 58098DCDF5B8CFE2 CRC64;

```

```

Query Match 75.4%; Score 264; DB 4; Length 330;
Best Local Similarity 100.0%; Pred. No. 4.3e-255;
Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 21 ILKSPGVSFRNGNMPDGERIPDVALSMGVSVEDLSMPGLAVGMLFRRPRATVWV 80
DB 1 ILKSPGVSFRNGNMPDGERIPDVALSMGVSVEDLSMPGLAVGMLFRRPRATVWV 80
QY 81 KGKVKLALPPGVSISYPLENAVPSLDSVANSIHSLFSEETPVVQLADSEERYVMGKA 140
DB 81 KGKVKLALPPGVSISYPLENAVPSLDSVANSIHSLFSEETPVVQLADSEERYVMGKA 140
QY 141 NSVFEDLSYTLQOLNRRIFOENSVLSLPLNSLNNEVDLFLSELQVLDHDISSLSRH 200
DB 141 NSVFEDLSYTLQOLNRRIFOENSVLSLPLNSLNNEVDLFLSELQVLDHDISSLSRH 200
QY 201 KHLAKHSDPDISLELAGLDEIGKRYGDSSEFRDASKITLVDALQRFADDMYSILYGNV 260
DB 181 KHLAKHSDPDISLELAGLDEIGKRYGDSSEFRDASKITLVDALQRFADDMYSILYGNV 240
QY 261 VELVTKSEFDSILRKTRITILEAK 284
DB 241 VELVTKSEFDSILRKTRITILEAK 264

```

```

RESULT 3
Q9RNX0 PRELIMINARY: PRT: 343 AA.
AC Q9RNX0:
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 16, Last annotation update)
DE IRP1B.
GN IRP1B.
OS Corynebacterium diphtheriae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Corynebacteriaceae;
OC Corynebacterium.
OX NCBI_TaxID=1717;

```

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C7;  
 RA Qian Y., Holmes R.K.;  
 RT "ABC transporters involved in Iron uptake in Corynebacterium  
 diptheriae";  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF176902; AAD53126.1; -  
 DR InterPro: IPR000522; -  
 DR Pfam: PF01032; Pfam family: 1.  
 SQ SEQUENCE 343 AA; 35017 MW; 332E5E5D13F4D52 CRC64;

Query Match 2.6%; Score 9; DB 2; Length 343;  
 Best Local Similarity 100.0%; Pred. No. 2;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 VEVVLLALV 11  
 |||||  
 DB 21 VEVVLLALV 29

RESULT 4  
 Q9S2H2 PRELIMINARY; PRT; 55 AA.  
 AC Q9S2H2;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DE HYPOTHETICAL 6.0 KDA PROTEIN.  
 GN SC19A.09.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxId=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Seeger K., Harris D.;  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 DE [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RC MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 the 8 Mb Streptomyces coelicolor A3(2) chromosome";  
 RL MCL. Microbiol. 21:77-96(1996).  
 DR EMBL: AL096852; CAB50990.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 55 AA; 5971 MW; C5C2BECDB0A253F9 CRC64;

Query Match 2.3%; Score 8; DB 2; Length 55;  
 Best Local Similarity 100.0%; Pred. No. 3.8;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 5 VVLLALVA 12  
 |||||  
 DB 12 VVLLALVA 19

RESULT 5  
 Q9HRH6 PRELIMINARY; PRT; 156 AA.  
 ID Q9HRH6;  
 AC Q9HRH6;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DR VNG0692C.  
 GN VNG0692C.  
 OS Halobacterium sp. (strain NRC-1).  
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
 OC Halobacterium  
 OX NCBI\_TaxId=64091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20504483; PubMed=11016950;  
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,  
 RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Shroana J.,  
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welt R., Goo Y.A.,  
 RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Spudis J.L., Jung K.-H.,  
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dennis P.P., Omer A.D.,  
 RA "Genome sequence of Halobacterium species NRC-1";  
 RT Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 DR EMBL: AE05015; AAG19182.1; -  
 SQ SEQUENCE 156 AA; 16056 MW; BD758B7D65FC2F45 CRC64;

Query Match 2.3%; Score 8; DB 1; Length 156;  
 Best Local Similarity 100.0%; Pred. No. 9.8;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 7 LLLAVGV 14  
 |||||  
 DB 63 LLLAVGV 70

RESULT 6  
 Q9H276 PRELIMINARY; PRT; 239 AA.  
 AC Q9H276;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE SEROLOGICALLY DEFINED BREAST CANCER ANTIGEN NY-BR-84 (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BREAST;  
 RA Scanlan M.J., Gout I., Stockert E., Gure A.O., Jaeger D., Chen Y.-T.,  
 RA Old L.J.;  
 RT "Humoral Immunity to Human Breast Cancer: Antigen Definition and  
 Quantitative Analysis of mRNA Expression";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF308298; AAC48265.1; -  
 FT NON\_TER  
 SQ SEQUENCE 239 AA; 27120 MW; 6F9A0BDA233669A5 CRC64;

Query Match 2.3%; Score 8; DB 4; Length 239;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 181 LFLFSELD 188  
 |||||  
 DB 49 LFLFSELD 56

RESULT 7  
 Q9VHG4 PRELIMINARY; PRT; 320 AA.  
 ID Q9VHG4;  
 AC Q9VHG4;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)  
 DE 01-MAY-2000 (TREMBLrel. 13, last annotation update)  
 DE CG8444 PROTEIN.  
 GN CG8444.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 AD Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 AMannatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 Amaniatis P.G., Scherer S.E., Richards S., Ashburner M., Henderson S.N.,  
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 Brandon R.C., Rogers Y.-H.C., Blake J.R., Champe M., Pfeiffer B.D.,  
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 Abail J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 Beeson K.Y., Benos P.V., Beremand B.P., Bhandari D., Bolshakov S.,  
 Borovaya D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 Kimmel B.E., Kodira C.D., Kratz C., Kravitz S., Kul D., Lai Z.,  
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 Palazzolo M., Pittman K.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 "the genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AEO03682; AAF94350.1; -;  
 DR FLYBase: FBgn003671; CG8444.  
 SQ SEQUENCE 320 AA; 34421 MW; 401BCBAAA4FBAT5 CRC64;

Query Match 2.3%; Score 8; DB 5; Length 320;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CN R03E1.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelioidinae; Caenorhabditis.  
 NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA McMurray A.;  
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,  
 Bonfield J., Burton J., Connell M., Cosey T., Cooper J., Coulson A.,  
 Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 Parsons J., Percy C., Sonhammer E., Roopra A., Saunders D., Showkneen R.,  
 Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 Thierly-Mieg J., Thomas K., Vaudin H., Vaughan K., Waterston R.,  
 Watson A., Weinstock L., Wilkinson-Sproat J., Wombliman P.,  
 RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans.";  
 RT Nature 368:32-38(1994).  
 RL Nature 368:32-38(1994).  
 DR EMBL: 292837; CAB07401.1; -;  
 SQ SEQUENCE 324 AA; 34217 MW; DOB3D186CF861C6E CRC64;

Query Match 2.3%; Score 8; DB 5; Length 324;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 336 DSIIYRMT 343  
 DB 310 DSIIYRMT 317  
 ID 025319 PRELIMINARY; PRT; 340 AA.  
 AC 025319;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)  
 DE CATHEPSIN B-LIKE ENZYME.  
 GN LMCPC.  
 OS Leishmania mexicana.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 NCBI\_TaxID=5665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MNY/BZ/62/M379;  
 RX MEDLINE=94187801; PubMed=8139620;  
 RA Bart G., Coombs G.H., Mottam J.C.;  
 "Isolation of lmcpc, a gene encoding a Leishmania mexicana  
 cathepsin-B-like cysteine proteinase.";  
 RL Mol. Biochem. Parasitol. 73:271-274(1995).  
 DR EMBL: 248599; CA88490.1; -;  
 DR HSP: P07858; IHDC.  
 DR MEROPS: C01.060; -;  
 DR InterPro: IPR000169; -;  
 DR InterPro: IPR000668; -;  
 DR Pfam: PF00112; Peptidase\_C1.1.  
 DR PRINTS: PR00705; PAPAIN.  
 DR PROSITE: PS00139; THIOL\_PROTEASE\_CYS; UNKNOWN\_1.

DR PROSITE; PS00639; THIOI\_PROTEASE\_HIS. 1.  
 KW Hydrolyase; Thiol protease.  
 FT CHAIN 97 340 CATHRPSIN B-LIKE CYSTEINE PROTEINASE.  
 SQ SEQUENCE 340 AA; 37200 MW; 15680C7862E2C36 CRC64;

Query Match 2.3%; Score 8; DB 5; Length 340;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 AVEPVLLA 9  
 |||||  
 DB 13 AVEPVLLA 20

RESULT 10  
 O9PIL3 PRELIMINARY; PRT; 346 AA.

AC O9PIL3;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DE PRO0989.

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 CX NCBI\_TaxID=9606;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LIVER;  
 RA Zhang C., Yu Y., Zhang S., Wei H., Zhou G., Ouyang S., Luo L., Bi J.,  
 RA Liu M., He F.;

RT "Functional prediction of the coding sequences of 121 new genes  
 deduced by analysis of cDNA clones from human fetal liver.";  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF161614; AAF71038.1; -

SO SEQUENCE 346 AA; 39238 MW; A042221C998834BF CRC64;

OY 181 LFLFSLQ 188  
 |||||  
 DB 3 LFLFSLQ 10

Query Match 2.3%; Score 8; DB 4; Length 346;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 11  
 O9Y282 PRELIMINARY; PRT; 383 AA.

AC O9Y282;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE CGI-54 PROTEIN (DJ47704.2) (CGI-54).  
 GN DJ47704.2.

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 CX NCBI\_TaxID=9606;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LIVER;  
 RA Lin W.-C.;

RT "Comparative gene cloning: Identification of novel human genes with  
 Caenorhabditis elegans proteome as template.";  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF077030; AAD27763.1; -

SO SEQUENCE 383 AA; 43222 MW; 8370FB3067AD95CE CRC64;

RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Brown A.;

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF151812; AAD34049.1; -  
 DR EMBL: AF077030; AAD27763.1; -  
 DR EMBL: AL121586; CAB89412.1; -

SO SEQUENCE 383 AA; 43222 MW; 8370FB3067AD95CE CRC64;

OY 181 LFLFSLQ 188  
 |||||  
 DB 40 LFLFSLQ 47

Query Match 2.3%; Score 8; DB 4; Length 383;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 12  
 O17498 PRELIMINARY; PRT; 438 AA.

AC O17498;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE PHENYLALANINE HYDROXYLASE (EC 1.14.16.1).  
 GN PAH.

OS Branchiostoma floridae (Florida lancelet) (Amphioxus).  
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;  
 OC Branchiostoma.  
 CX NCBI\_TaxID=7739;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LIVER;  
 RA Patton S.J., Luke G.N., Holland P.W.;

RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ001677; CAA04917.1; -  
 DR HSSP: P04176; 1PHZ.

DR InterPro: IPR001273; -  
 DR InterPro: IPR002119; -  
 DR InterPro: IPR002912; -  
 DR Pfam: PF00351; Biotin\_H; 2.  
 DR Pfam: PF01842; ACT; 1.

DR PROSITE; PS00046; HISTONE\_H2A; UNKNOWN\_1.  
 DR Oxidoreductase.  
 SO SEQUENCE 438 AA; 49958 MW; 1619297DBDBF5EE7 CRC64;

OY 239 ILVDALQK 246  
 |||||  
 DB 428 ILVDALQK 435

Query Match 2.3%; Score 8; DB 5; Length 438;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 13  
 O9FWM6 PRELIMINARY; PRT; 468 AA.

AC O9FWM6;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE T28K15.10 PROTEIN.  
 GN T28K15.10.

OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 CX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,  
 RA Altifi H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,  
 RA Buehler E., Chou Q., Chin C., Chou J., Choi E., Gonzalez A.,  
 RA Hong B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,  
 RA Lenz C., Liu A., Liu S., Mukharly N., Pham P., Sakano H., Shin P.,  
 RA Toriumi M., Vayshberg M., Yu G., Ecker J., Theologos A., Davis R.W.,  
 RA Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC022522; MAG1257.1; -;  
 SQ SEQUENCE 468 AA; 53271 MW; SDIFR79463CDAF20 CRC64;

Query Match 2.3%; Score 8; DB 10; Length 468;  
 Best Local Similarity 100.0%; Pred. No. 27;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 GSVFRNG 33  
 DB 261 GSVFRNG 268

RESULT 14

ID 09LDK2 PRELIMINARY; PRT; 486 AA.

AC 09LDK2: PRELIMINARY; PRT; 486 AA.  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE P0512G09.5 PROTEIN (SIMILAR TO ARABIDOPSIS THALIANA CHROMOSOME 2  
 DE SECTION 175 OF 255).  
 GN P0512G09.5.

OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzae;  
 OC Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 RT clone: P0512G09.5";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 RT clone: P0695A04.1";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF002836; BAB07923.1; -;  
 DR EMBL: AF002816; BAB03416.1; -;  
 DR InterPro: IPR000477; -;  
 DR Pfam: PF00078; ITC; 1.  
 SQ SEQUENCE 486 AA; 57210 MW; 8F191E5B319D224 CRC64;

Query Match 2.3%; Score 8; DB 10; Length 486;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 162 NSVLSLP 169  
 DB 95 NSVLSLP 102

RESULT 15

ID 09FUT8 PRELIMINARY; PRT; 610 AA.

AC 09FUT8: PRELIMINARY; PRT; 610 AA.  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE GBLAAD39327.1.

OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;  
 RX MEDLINE=99087489; PubMed=9872454;  
 RA Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.  
 RT Sequence features of the regions of 1,013,767 bp covered by sixteen  
 RT physically assigned P1 and TAC clones.";  
 RL DNA Res. 5:297-308 (1998).  
 DR EMBL: AB015469; BAB11508.1; -;  
 SQ SEQUENCE 610 AA; 66191 MW; F6BFB6896FBADE2 CRC64;

Query Match 2.3%; Score 8; DB 10; Length 610;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 167 SUPLSLS 174  
 DB 78 SUPLSLS 85

RESULT 16

ID 09LDS6 PRELIMINARY; PRT; 656 AA.

AC 09LDS6: PRELIMINARY; PRT; 656 AA.  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE SERINE/THREONINE KINASE-LIKE PROTEIN.  
 GN A74G1480.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA ED Arabidopsis sequencing project;  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Leonard N., Quail M., Harris B., Rajandream M.A., Barrell B.G.,  
 RA Lewis H.W., Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL: AL050399; CAB82153.1; -;  
 DR EMBL: AL161532; CAB78191.1; -;  
 DR InterPro: IPR000719; -;  
 DR InterPro: IPR002290; -;  
 DR InterPro: IPR002902; -;  
 DR Pfam: PF00069; Pkinase; 1.  
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE: PS50011; PROTEIN KINASE DOM; 1.  
 DR PROSITE: PS00106; PROTEIN KINASE\_ST; 1.  
 DR SMART: SM00220; S\_TKc; 1.  
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.

DR ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
 SQ SEQUENCE 656 AA; 73695 MW; 61099983E01F70CA CRC64;

Query Match 2.3%; Score 8; DB 10; Length 656;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 FVVLALV 11  
 DB 4 FVVLALV 11

DB 273 FVLLALV 280

RESULT 17

09M04 PRELIMINARY; PRT: 695 AA.

AC 09M04; (TREMblrel. 15, Created)

DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)

DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)

DE NADH DEHYDROGENASE SUBUNIT (FRAGMENT).

GN NDHF.

OS Solanum nltidum.

OC Chloroplast.

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons: core eudicots; Asteridae; euasterids I;

OC Solanales; Solanales; Solanum.

OC NCBI\_TaxID=115668;

OX NCBI\_TaxID=115668;

RN [1]

RP SEQUENCE FROM N.A.

RA Bohs L., Olmstead R.G.;

RT A reassessment of Normonia and Triquera (Solanaceae).";

RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF224075; AAF35882.1;

DR InterPro: IPR001064;

DR InterPro: IPR001516;

DR InterPro: IPR001750;

DR InterPro: IPR002128;

DR Pfam: PF00361; oxidored\_q1; 1.

DR Pfam: PF00662; oxidored\_q1\_N; 1.

DR Pfam: PF01010; oxidored\_q1\_C; 1.

DR PROSITE: PS00225; CRYSTALLIN\_BETAGAMMA; UNKNOWN\_1.

KW Chloroplast.

FT NON\_TER 1 1

FT NON\_TER 695 695

SO SEQUENCE 695 AA; 78384 MW; BD41140DDEB309 CRC64;

Query Match 2.3%; Score 8; DB 8; Length 695;

Best Local Similarity 100.0%; Pred. No. 39;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 NNEVDLF 183

DB 210 NNEVDLF 217

RESULT 18

09V97 PRELIMINARY; PRT: 890 AA.

AC 09V97; (TREMblrel. 13, Created)

DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)

DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)

DE CG17149 PROTEIN.

GN CG17149.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Anthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephyroidea; Drosophilidae; Drosophila.

OC NCBI\_TaxID=7227;

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN-BERKELEY;

RC MEDLINE-20196006; PubMed-10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brother P.,

RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Durkin B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Poser C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang X., Lin X.,

RA Liu X., Mottel B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merklov G., Mishina N.V., Mobarry C., Morris J., Moshel A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclebo J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhu G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT The genome sequence of Drosophila melanogaster.;

RL Science 287:2185-2195(2000).

DR EMBL: AE003514; AAF49051.1;

DR HSSP: 064411; 1837

DR Flybase: FBgn0036955; CG17149.

DR InterPro: IPR000205;

DR InterPro: IPR000759;

DR InterPro: IPR002937;

DR Pfam: PF01593; Amino-oxidase; 1.

DR PRINTS: PR00419; ADXRPTASE.

SO SEQUENCE 890 AA; 98388 MW; 47B1CA7F7404827 CRC64;

Query Match 2.3%; Score 8; DB 5; Length 890;

Best Local Similarity 100.0%; Pred. No. 49;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLLALVAG 13

DB 703 VLLALVAG 710

RESULT 19

09ZVW0 PRELIMINARY; PRT: 1413 AA.

AC 09ZVW0; (TREMblrel. 10, Created)

DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)

DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)

DE PUTATIVE RETROTRANSPOSON.

GN T4E14.12.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons: core eudicots; Rosidae; eurosids II;

OC Brassicales; Brassicaceae; Arabidopsis.

OC NCBI\_TaxID=3702;

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN-CV. COLOMBIA;

RC Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,

RA Shen M., Rongling C.M., Fraser C.M., Somerville C.R., Venter J.C.,

RT Arabidopsis thaliana chromosome II BAC T4E14 genomic sequence.;

RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO ZN-FINGER CCHC TYPE FAMILY.

DR EMBL: AC005171; AAC67205.1;

DR InterPro: IPR001584; -;  
 DR InterPro: IPR001878; -;  
 DR Pfam: PF00098; zf-CCHC; 1.  
 DR SMART: SM00665; rve; 1.  
 DR SMART: SM00343; znf\_C2HC; 1.  
 DR Zinc-finger.  
 KW ZINC\_FINGER.  
 SEQUENCE 1413 AA; 157949 MW; C8450A69DC1241A9 CRC64;

Query Match 2.3%; Score 8; DB 10; Length 1413;  
 Best Local Similarity 100.0%; Pred. No. 75;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 SVLSSLP 170  
 |||||  
 DB 523 SVLSSLP 530

RESULT 20  
 Q91VQ2 PRELIMINARY; PRT; 1491 AA.

AC Q91VQ2;  
 DT 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
 DE RETROELEMENT POLYPROTEIN-LIKE.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COLUMBIA;  
 RX MEDLINE=20181125; PubMed=10718197;  
 RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence  
 RT features of the regions of 3,076,755 bp covered by sixty pl and TAC  
 RT clones.";  
 RL DNA Res. 7:31-63(2000).  
 DR EMBL: AB018120; BAA9787.1; -;  
 DR InterPro: IPR001584; -;  
 DR InterPro: IPR001878; -;  
 DR Pfam: PF00098; zf-CCHC; 1.  
 DR Pfam: PF00665; rve; 1.  
 KW Polyprotein.  
 SQ SEQUENCE 1491 AA; 166720 MW; C0998D6872BC5AA7 CRC64;

Query Match 2.3%; Score 8; DB 10; Length 1491;  
 Best Local Similarity 100.0%; Pred. No. 78;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 SVLSSLP 170  
 |||||  
 DB 523 SVLSSLP 530

RESULT 21  
 Q92P04 PRELIMINARY; PRT; 1501 AA.

AC Q92P04;  
 DT 01-MAY-1999 (TReMBLrel. 10, Created)  
 DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
 DE PUTATIVE POLYPROTEIN WITH A ZN-FINGER CCHC TYPE DOMAIN  
 DE (PROSITE:OPC05158) AND A DDE INTEGRASE SIGNATURE MOTIF.  
 DE F9B22.5.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COLUMBIA;  
 RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., Vanaken S.E.,  
 RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.,  
 RA Carreir A.J., Greasy T.H., Buell C.R., Town C.D., Nierman W.C.,  
 RA Fraser C.M., Venter J.C.;  
 RT "Arabidopsis thaliana chromosome II BAC F9B22 genomic sequence.";  
 RT Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO ZN-FINGER CCHC TYPE FAMILY.  
 DR EMBL: AC006528; AB019784.1; -;  
 DR InterPro: IPR001584; -;  
 DR InterPro: IPR001878; -;  
 DR Pfam: PF00098; zf-CCHC; 1.  
 DR Pfam: PF00665; rve; 1.  
 DR SMART: SM00343; znf\_C2HC; 1.  
 KW Polyprotein; Zinc-finger.  
 SQ SEQUENCE 1501 AA; 167438 MW; 166F995793164239 CRC64;

Query Match 2.3%; Score 8; DB 10; Length 1501;  
 Best Local Similarity 100.0%; Pred. No. 79;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 SVLSSLP 170  
 |||||  
 DB 540 SVLSSLP 547

RESULT 22  
 Q63375 PRELIMINARY; PRT; 1715 AA.

AC Q63375;  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
 DE NEUREXIN II-ALPHA-A-PRECURSOR.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92320296; PubMed=1621094.  
 RA Ushakov Y.A., Petrenko A.G., Geppert M., Sudhof T.C.;  
 RT "Neurexins: synaptic cell surface proteins related to the alpha-  
 RT latrotoxin receptor and laminin.";  
 RL Science 257:50-56(1992).  
 DR EMBL: M96376; AAA41707.1; -;  
 DR InterPro: IPR000561; -;  
 DR InterPro: IPR001791; -;  
 DR InterPro: IPR003585; -;  
 DR Pfam: PF00008; EGF; 3.  
 DR Pfam: PF00054; Laminin\_G; 5.  
 DR SMART: SM00294; 4.1m; 1.  
 KW SIGNAL.  
 FT SIGNAL 1 27 POTENTIAL.  
 FT CHAIN 28 1715 NEUREXIN II-ALPHA-B.  
 FT CHAIN 28 1715 NEUREXIN II-ALPHA-A.  
 SQ SEQUENCE 1715 AA; 185282 MW; 59FBF18661F3DB15 CRC64;

Query Match 2.3%; Score 8; DB 11; Length 1715;  
 Best Local Similarity 100.0%; Pred. No. 89;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LLALVAGV 14  
 |||||  
 DB 19 LLALVAGV 26

RESULT 23



```

063374 ID 063374 PRELIMINARY; PRT: 1728 AA.
AC 063374;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE NEUREXIN II-ALPHA-B PRECURSOR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92320296; PubMed=1621094;
RA Ushekarov Y.A., Petrenko A.G., Geppert M., Sudhof T.C.;
RT "Neurexins: synaptic cell surface proteins related to the alpha-
RT latrotoxin receptor and laminin.";
RT Science 257:50-56(1992).
DR EMBL: M8376; AAA41706.1; -.
DR InterPro: IPR000561; -.
DR InterPro: IPR001791; -.
DR Pfam: PF00008; EGF_3.
DR Pfam: PF00054; Laminin_G: 5.
DR SMART: SM00282; Lamin; 1.
KW Signal.
FT SIGNAL. 1 27 POTENTIAL.
FT CHAIN 28 1728 NEUREXIN II-ALPHA-B.
FT CHAIN 28 1715 NEUREXIN II-ALPHA-A.
SQ SEQUENCE 1728 AA; 186626 MW; 52EA5FA8BC585784 CRC64;

Query Match 2.3%; Score 8; DB 11; Length 1728;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LIALVAGV 14
DB 19 LIALVAGV 26

RESULT 24
091179 ID 091179 PRELIMINARY; PRT: 5149 AA.
AC 091179;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PROBABLE NON-RIBOSOMAL PEPTIDE SYNTHETASE.
GN PA2402.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.O., Lagrou M.,
RA Gader R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RT Nature 406:959-964(2000).
RX EMBL: AB004667; AAC05790.1; -.
DR InterPro: IPR000255; -.
DR InterPro: IPR000276; -.
DR InterPro: IPR000577; -.
DR InterPro: IPR000873; -.
DR InterPro: IPR001242; -.
DR Pfam: PF00501; AMP-binding; 4.

```

```

DR Pfam: PF00550; pp-binding; 4.
DR Pfam: PF00668; Condensation; 6.
DR PROSITE: PS0075; ACP_DOMAIN; 4.
DR PROSITE: PS00455; AMP_BINDING; 4.
DR PROSITE: PS00445; FGK_KINASES_2; UNKNOWN_1.
DR PROSITE: PS00237; G_PROTEIN_RECEPTOR; UNKNOWN_3.
DR PROSITE: PS00012; PHOSPHOPANTHETINE; UNKNOWN_3.
SQ SEQUENCE 5149 AA; 569179 MW; 2057587618921CA0 CRC64;

Query Match 2.3%; Score 8; DB 2; Length 5149;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 VLSUPLN 171
DB 2462 VLSUPLN 2469

RESULT 25
0903E5 ID 0903E5 PRELIMINARY; PRT: 67 AA.
AC 0903E5;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE F53H2.2 PROTEIN.
GN F53H2.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Lloyd C.R.;
RL submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=99069613; PubMed=9851916;
RP SEQUENCE FROM N.A.
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL: Z82270; CAB63204.1; -.
SQ SEQUENCE 67 AA; 7324 MW; D80E860A3793D8BA CRC64;

Query Match 2.0%; Score 7; DB 5; Length 67;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LIALVAG 13
DB 9 LIALVAG 15

RESULT 26
035298 ID 035298 PRELIMINARY; PRT: 73 AA.
AC 035298;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-JAN-1999 (TREMBLrel. 09, Last annotation update)
DE ORF73.
OS Oryza sativa (Rice).
OC Mitochondrion.
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzae;
OX Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MTG-5A.

```

RA MEDLINE-94363737; PubMed-7915966;  
 RA Arai H., Sakamoto M., Shimada H., Shinjo C., Fujimura T.;  
 RT "A unique sequence located downstream from the rice mitochondrial atp6  
 RT may cause male sterility";  
 RL Curr. Genet. 25:52-58(1993).  
 RL EMBL: D14339; BAA03272.1; -.  
 DR Mitochondrion.  
 KW Mitochondrion.  
 SQ SEQUENCE 73 AA; 8437 MW; 73937D5C7F3AF8C CRC64;

Query Match 2.0%; Score 7; DB 8; Length 73;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 102 VPPSLDS 108  
 DB 50 VPPSLDS 56

RESULT 27  
 ID Q9NAJ6 PRELIMINARY; PRT; 74 AA.  
 AC Q9NAJ6;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE Y38E10A.15 PROTEIN.  
 GN Y38E10A.15.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 NC NCB1\_Taxid-6239;  
 RX NCB1\_Taxid-6239;  
 RN SEQUENCE FROM N.A.  
 RP Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RL [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-99069613; PubMed-9851916;  
 RA none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 RT investigating biology.";  
 RL Science 282:2012-2018(1998).  
 DR EMBL: AL110484; CAB54406.1; -.  
 SQ SEQUENCE 74 AA; 8187 MW; 57DC0FE43F51E3CE CRC64;

Query Match 2.0%; Score 7; DB 5; Length 74;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 VFVYLLA 9  
 DB 10 VFVYLLA 16

RESULT 28  
 ID Q9NMP7 PRELIMINARY; PRT; 126 AA.  
 AC Q9NMP7;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE CCNA FLJ20700 FIS, CLONE KAI1A2250.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 NC NCB1\_Taxid-9606;  
 RX NCB1\_Taxid-9606;  
 RN SEQUENCE FROM N.A.  
 RP TISSUE-ILKLV MUCOSA.  
 RA Tanigami A., Fujiwara T., Ono T., Yamada K., Fujii Y., Ozaki K.,  
 Hirao M., Ohmori Y., Ota T., Suzuki Y., Ohashi M., Nishi T.,

RA Shibahara T., Tanaka T., Nakamura Y., Isogai T., Sugano S.;  
 RT "MEDD human cDNA sequencing project.";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK000707; BAA91332.1; -.  
 SQ SEQUENCE 126 AA; 13762 MW; E430EE2539536879 CRC64;

Query Match 2.0%; Score 7; DB 4; Length 126;  
 Best Local Similarity 100.0%; Pred. No. 81;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 163 SVLSSLP 169  
 DB 105 SVLSSLP 111

RESULT 29  
 ID Q9F3C8 PRELIMINARY; PRT; 136 AA.  
 AC Q9F3C8;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE PUTATIVE SULFATASE (FRAGMENT).  
 GN SC5F1.01.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycetaceae; Streptomyces.  
 NC NCB1\_Taxid-1902;  
 RX NCB1\_Taxid-1902;  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-A3(2);  
 RL Saunders D.C., Harris D.;  
 RA Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN-A3(2);  
 RC MEDLINE-97000351; PubMed-8843436;  
 RA Redenbach M., Kieser H.M., Denapite D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmid and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL: AL450165; CAC16427.1; -.  
 FT NON\_TER 1  
 SQ SEQUENCE 136 AA; 15875 MW; 0D9462442EC5C866 CRC64;

Query Match 2.0%; Score 7; DB 2; Length 136;  
 Best Local Similarity 100.0%; Pred. No. 86;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 214 LELAGLD 220  
 DB 2 LELAGLD 8

RESULT 30  
 ID O52521 PRELIMINARY; PRT; 137 AA.  
 AC O52521;  
 DT 01-JUN-1998 (TREMBlrel. 06, Created)  
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE CTTCHROME C OXIDASE CCOO SUBUNIT (FRAGMENT).  
 GN FIXO.  
 OS Rhizobium fredii (Sinorhizobium fredii).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Rhizobiaceae; Sinorhizobium.  
 OX NCBI\_TaxID=380;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RT19;  
 RA Bian X.L., Sun H.H., Gao W.M., Yang S.S.;  
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF039956; AAC02673.1; -  
 DR InterPro: IPR003468; -  
 DR Pfam: PF02433; Pfam: 1.  
 FT NON\_TER 137 137  
 SQ SEQUENCE 137 AA; 15751 MW; 4E3A62BA684C35DC CRC64;

Query Match 2.0%; Score 7; DB 2; Length 137;  
 Best Local Similarity 100.0%; Pred. No. 87;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 9 ALVAVG 15  
 DB 7 ALVAVG 13

RESULT 31  
 O99201  
 ID 099201; PRELIMINARY; PRT; 137 AA.  
 AC 099201;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DE 01-MAR-2001 (TREMBLrel. 13, Last annotation update)  
 DE HYPOTHEICAL 15.8 KDA PROTEIN IN TOC1.1 TRANSPOSON.  
 GN ORF31.  
 OS Chlamydomonas reinhardtii.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 OC Chlamydomonadaceae; Chlamydomonas.  
 OX NCBI\_TaxID=3053;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FUD4;  
 RX MEDLINE=91232906; PubMed=1851555;  
 RA Day A., Rochaix J.D.;  
 RT "A transposon with an unusual LTR arrangement from Chlamydomonas  
 RL reinhardtii contains an internal tandem array of 76 bp repeats."  
 RL Nucleic Acids Res. 19;1259-1266(1991).  
 DR EMBL: X56231; CAA39687.1; -  
 DR PIR: S14025; S14025.  
 KW Hypothetical protein; Transposable element.  
 SQ SEQUENCE 137 AA; 15781 MW; 463630E155E5E942 CRC64;

Query Match 2.0%; Score 7; DB 10; Length 137;  
 Best Local Similarity 100.0%; Pred. No. 87;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 124 VLQAPS 130  
 DB 17 VLQAPS 23

RESULT 32  
 O9N2L6  
 ID 09N2L6; PRELIMINARY; PRT; 141 AA.  
 AC 09N2L6;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE W08E12.4 OR W08E12.5 PROTEIN.  
 GN W08E12.4 OR W08E12.5.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodermatidae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL NZ;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RL investigating biology. The C. elegans Sequencing Consortium."  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL NZ;  
 RA Miller N., Magi L., Goela D., Andrews S.;  
 RT "The sequence of C. elegans cosmid W08E12."  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL NZ;  
 RA Waterston R.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC006696; AAF39987.1; -  
 DR EMBL: AC006696; AAF39986.1; -  
 DR InterPro: IPR000564; -  
 DR PROSITE: PS00197; 2FEZS\_FERREDOXIN; UNKNOWN 1.  
 SQ SEQUENCE 141 AA; 14247 MW; 2E19E5DC375E9BD CRC64;

Query Match 2.0%; Score 7; DB 5; Length 141;  
 Best Local Similarity 100.0%; Pred. No. 89;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 7 LALVAG 13  
 DB 9 LALVAG 15

RESULT 33  
 O30276  
 ID 030276; PRELIMINARY; PRT; 148 AA.  
 AC 030276;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE IRON-DEPENDENT REPRESSOR.  
 GN AF2395.  
 OS Archaeoglobus fulgidus.  
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;  
 OC Archaeoglobus.  
 OX NCBI\_TaxID=2234;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
 RX MEDLINE=98049343; PubMed=9389475;  
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriakides N.C.,  
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
 RA Peterson S., Reich C.I., McNeil L.K., Bauger J.H., Glodek A., Zhou L.,  
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
 RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,  
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the hyperthermophilic, sulphate-  
 RL reducing archaeon Archaeoglobus fulgidus."  
 RL Nature 390:364-370(1997).  
 DR EMBL: AE001111; AAB91276.1; -  
 DR HSSP: Q50495; IBB.  
 DR TIGR: AF2395; -  
 DR InterPro: IPR001367; -  
 DR Pfam: PF01325; Fe\_dep\_repress; 1.  
 DR SMART: SM00529; HTH\_DTXR; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 148 AA; 17304 MW; 1E39FC73D0C86C88 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 148;  
 Best Local Similarity 100.0%; Pred. No. 93;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 TLRLRN 156  
 |||||  
 Db 109 TLRLRN 115

RESULT 34  
 ID Q9HVX4 PRELIMINARY; PRT; 148 AA.

AC Q9HVX4; 01-MAR-2001 (TREMblrel. 16, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
 DE 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
 DE HYPOTHETICAL PROTEIN PA4441.  
 GN PA4441.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 CC Pseudomonas.  
 CX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PA01;  
 RX MEDLINE-20437337; PubMed-10984043;  
 RA Stever C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltz L., Tolentino E., Westbrock-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lartig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Lu Z., Paulsen I.T.,  
 RA Ralzer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 RT opportunistic pathogen";  
 RL Nature 406:959-964(2000).  
 KW EMBL; AE004858; AAG07829.1; -.  
 DR Hypothetical protein.  
 SQ SEQUENCE 148 AA; 16418 MW; 4DD9E4B217BD2B8 CRC64;

Query Match 2.0%; Score 7; DB 2; Length 148;  
 Best Local Similarity 100.0%; Pred. No. 93;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LALVAV 14  
 |||||  
 Db 13 LALVAV 19

RESULT 35  
 ID Q9MLQ9 PRELIMINARY; PRT; 160 AA.

AC Q9MLQ9; 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DE 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
 DE NADH DEHYDROGENASE SUBUNIT 2 (FRAGMENT).  
 GN ND2.  
 OS Gymnotus carapo.  
 OC Gymnotus.  
 OC Mitochondrion.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
 OC Gymnotiformes; Gymnotidae; Gymnotus.  
 CX NCBI\_TaxID=94172;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-4-G56;  
 RA Fernandes-Matoli F.M.C., Matoli S.R., Almeida-Toledo L.F.;  
 RT "The coalescent process in Gymnotus carapo (Pisces: Gymnotidae)  
 RT inferred from ND2 and COI partial sequences";  
 RL Submitted (DEC-1999) to the EMBL/Genbank/DDBJ databases.

DR EMBL; AF214003; AAF35242.1; -  
 DR InterPro: IPR001750; -.  
 DR Pfam: PF00361; oxidored\_g1; 1.  
 KW Mitochondrion.  
 FT NON\_TER 1 1  
 FT NON\_TER 160 160  
 SQ SEQUENCE 160 AA; 17789 MW; 2772A6D6F0998175 CRC64;

Query Match 2.0%; Score 7; DB 8; Length 160;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 VLQLAPS 130  
 |||||  
 Db 5 VLQLAPS 11

RESULT 36  
 ID Q9RXD4 PRELIMINARY; PRT; 168 AA.

AC Q9RXD4; 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DE 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
 DE CONSERVED HYPOTHETICAL PROTEIN.  
 GN DR0380.  
 OS Deinococcus radiodurans.  
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
 CX NCBI\_TaxID=1299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RL;  
 RX MEDLINE-20036896; PubMed-10567266;  
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 RA Moritz K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
 RA Vamathevan J.J., Lam P., McDonald L., Uitterback T., Zalewski C.,  
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 RA Fraser C.M.;  
 RT "Genome sequence of the radioresistant bacterium Deinococcus  
 RT radiodurans RL";  
 RL Science 286:1571-1577(1999).  
 DR EMBL; AE001898; AAF0961.1; -.  
 DR TIGR; DR0380; -.  
 SQ SEQUENCE 168 AA; 18229 MW; 52ADA6D9CD36609 CRC64;

Query Match 2.0%; Score 7; DB 2; Length 168;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 IALALAV 321  
 |||||  
 Db 58 IALALAV 64

RESULT 37  
 ID Q9M6L6 PRELIMINARY; PRT; 172 AA.

AC Q9M6L6; 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DE 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
 DE HYPOTHETICAL 19.7 KDA PROTEIN (FRAGMENT).  
 GN Cajanus cajan (Pigeon pea).  
 OS Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
 OC Fabales; Fabaceae; Papilionoideae; Cajanus.  
 CX NCBI\_TaxID=3821;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Meyers B.C., Dickerman A.W., Michelmore R.W., Sivaramakrishnan S.,  
 RA Sobral B.W., Young N.D.:  
 RT "Plant disease resistance genes encode members of an ancient and  
 RT diverse protein family within the nucleotide-binding superfamily."  
 RL Submitted (Sep-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF186538; AAF6346.1; -  
 DR InterPro: IPR000767; -  
 DR Pfam: PF00931; NB-ARC.1.  
 DR PRINTS: PR00364; DISEASESERSIST.  
 KW Hypothetical protein.  
 FT NON\_TER 1  
 SQ SEQUENCE 172 AA; 19740 MW; 97C78EBB83A3ECB CRC64;

Query Match 2.0%; Score 7; DB 10; Length 172;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 277 TRITILEA 283  
 |||||  
 DB 41 TRITILEA 47

RESULT 38  
 031399 PRELIMINARY; PRT; 174 AA.

AC 031399; 01-NOV-1996 (TREMBLERL. 01, Created)  
 DT 01-NOV-1996 (TREMBLERL. 01, last sequence update)  
 DT 01-MAR-2001 (TREMBLERL. 16, last annotation update)  
 DE HMC B COMPLEX B-G SUBREGION (FRAGMENT).  
 OS Gallus gallus (Chicken).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 CC Gallus.  
 NX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86085308; PubMed=2826332;  
 RA Goto R., Miyada C.G., Young S., Wallace R.B., Abplanalp H.,  
 RA Bloom S.E., Briles W.E., Miller M.M.:  
 RT "Isolation of a cDNA clone from the B-G subregion of the chicken  
 RT histocompatibility (B) complex."  
 RL Immunogenetics 27:102-109(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91239571; PubMed=1903541;  
 RA Miller M.M., Goto R., Young S., Chirivella J., Hawke D., Miyada C.G.:  
 RT "Immunoglobulin variable-region-like domains of diverse sequence  
 RT within the major histocompatibility complex of the chicken."  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:4377-4381(1991).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96134001; PubMed=8550096;  
 RA Jarry S.I., Goto R.M., Briles W.E., Miller M.M.:  
 RT "Characterization of Mhc genes in a multigene family of ring-  
 RT necked pheasants."  
 RL Immunogenetics 43:125-135(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX Miyada C.G., Chirivella J., Goto R.M., Young S., Miller M.M.:  
 RL Submitted (Jul-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX  
 CC DOMAIN.  
 DR EMBL: U32560; AAC59846.1; -  
 DR InterPro: IPR003006; -  
 DR Pfam: PF00047; 1g.1  
 DR SMART: SM00410; 1g\_1like; 1.  
 KW MHC.  
 FT NON\_TER 1 1

FT NON\_TER 174 174  
 SQ SEQUENCE 174 AA; 19531 MW; 9A9061DAB5AD4637 CRC64;

Query Match 2.0%; Score 7; DB 7; Length 174;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 318 ALAVITT 324  
 |||||  
 DB 92 ALAVITT 98

RESULT 39  
 028534 PRELIMINARY; PRT; 179 AA.

AC 028534; 01-JAN-1998 (TREMBLERL. 05, Created)  
 DT 01-JAN-1998 (TREMBLERL. 05, last sequence update)  
 DT 01-JUN-2000 (TREMBLERL. 14, last annotation update)  
 DE CONSERVED HYPOTHETICAL PROTEIN.  
 GN AF1740.  
 OS Archaeoglobus fulgidus.  
 CC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;  
 CC Archaeoglobus.  
 NX NCBI\_TaxID=2234;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
 RX MEDLINE=9804943; PubMed=9389475;  
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
 RA Kirschman D.E., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriades N.C.,  
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
 RA Kirschman E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
 RA Peterson S., Reich C.I., McNeill L.K., Badger J.H., Glodde A., Zhou L.,  
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
 RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,  
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Venter J.C.:  
 RT "The complete genome sequence of the hyperthermophilic, sulphate-  
 RT reducing archaeon Archaeoglobus fulgidus."  
 RL Nature 390:364-370(1997).  
 DR EMBL: AE000982; AAB89505.1; -  
 DR TIGR: AF1740; -  
 DR InterPro: IPR002726; -  
 DR Pfam: PF01864; DUF46; 1.  
 DR ProDom: PD01601; -; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 179 AA; 19883 MW; 416F14FA4EC2B2AA CRC64;

Query Match 2.0%; Score 7; DB 1; Length 179;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 318 ALAVITT 324  
 |||||  
 DB 152 ALAVITT 158

RESULT 40  
 09Y8N1 PRELIMINARY; PRT; 184 AA.

AC 09Y8N1; 01-NOV-1999 (TREMBLERL. 12, Created)  
 DT 01-NOV-1999 (TREMBLERL. 12, last sequence update)  
 DT 01-MAR-2001 (TREMBLERL. 16, last annotation update)  
 DE HYPOTHETICAL 19.5 KDA PROTEIN APE2602.  
 GN APE2602.  
 OS Aeropyrum pernix.  
 CC Archaea; Crenarchaeota; Desulfurococcaceae; Desulfurococcaceae;

OC Aeropyrum  
 OX NCBI\_TaxID=56636;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K1:  
 RX MEDLINE=99310339; PubMed=10382966;  
 RA Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Halkawa Y.,  
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anka A., Kosugi H.,  
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,  
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,  
 RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,  
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.,  
 \*Complete genome sequence of an aerobic hyper-thermophilic  
 RT crenarchaeon, Aeropyrum pernix K1.\*;  
 RL DNA Res. 6:83-101(1999).  
 RM EMBL: AP000064; BAA81619.1;  
 KW hypothetical protein.  
 SQ SEQUENCE 184 AA; 19466 MW; F3BAF58E6F7D208C CRC64;

Query Match 2.0%; Score 7; DB 1; Length 184;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 317 LALAVIT 323  
 Db 15 LALAVIT 21

RESULT 41  
 ID 038468 PRELIMINARY; PRT: 195 AA.  
 AC 038468;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last annotation update)  
 DE URF 2.  
 OS Bacteriophage phi-H.  
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Podoviridae.  
 OX NCBI\_TaxID=10771;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP Schabel H., Palm P., Dick K., Grapp B.,  
 RL EMBO J. 3:1717-1722(1984).  
 DR EMBL: X00805; CAA25389.1;  
 SQ SEQUENCE 195 AA; 20984 MW; EC24140B85842BB3 CRC64;

Query Match 2.0%; Score 7; DB 9; Length 195;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 42 IPDVAA 48  
 Db 33 IPDVAA 39

RESULT 42  
 ID 023633 PRELIMINARY; PRT: 196 AA.  
 AC 023633;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE COSMID 2K84.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditicoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;

RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkhen R.,  
 RA Swaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Spoat J., Woldman P.,  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";  
 RL Nature 368:32-38(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Kirsten J.;  
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U23181; AAC48205.1;  
 SQ SEQUENCE 196 AA; 22442 MW; 1CC18435B3767CFA CRC64;

Query Match 2.0%; Score 7; DB 5; Length 196;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 84 NKLAAP 90  
 Db 162 NKLAAP 168

RESULT 43  
 ID 09RND0 PRELIMINARY; PRT: 202 AA.  
 AC 09RND0;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
 DE VIRB8 HOMOLOG.  
 OS Bartonella henselae (Rochalimaea henselae).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Bartonellaceae; Bartonella.  
 OX NCBI\_TaxID=38323;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN-HOUSTON-L1;  
 RA Padmalayam I., Baumstark B., Massung R.,  
 RT "The gene encoding the 17-kDa antigen of Bartonella henselae is  
 RT located within the virB operon.";  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF182718; AAF00946.1;  
 SQ SEQUENCE 202 AA; 23087 MW; 701DD6876A31B085 CRC64;

Query Match 2.0%; Score 7; DB 2; Length 202;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 315 IALAAV 321  
 Db 23 IALAAV 29

RESULT 44  
 ID 09LFF4 PRELIMINARY; PRT: 203 AA.  
 AC 09LFF4;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE HYPOTHETICAL 21.9 KDA PROTEIN.

GN PAP12\_370.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bloembergen H., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X., Quetier F.,  
 RA Salanoubat M.,  
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AL132966; CMB67674.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 203 AA; 21934 MW; 9E6AC99AFA8A8A65 CRC64;

Query Match 2.0%; Score 7; DB 10; Length 203;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 167 SLPLNSL 173  
 |||||  
 DB 70 SLPLNSL 76

RESULT 45  
 O67842 PRELIMINARY; PRT; 213 AA.  
 AC O67842;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE HYPOTHEICAL 25.2 KDA PROTEIN.  
 OS AO.2059.  
 GN Aquifex aeolicus.  
 OC Bacteria; Aquificales; Aquificaceae; Aquifex.  
 RN [1]  
 OX NCBI\_TaxID=63363;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VF5;  
 RA MEDLINE=98196666; PubMed=9537320;  
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
 RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aulley M., Huber R.,  
 RA Feldman R.A., Short J.M., Olson G.J., Swenson R.V.;  
 RT "The complete genome of the hyperthermophilic bacterium Aquifex  
 aeolicus.";  
 RL Nature 392:353-358(1998).  
 RT NATURE 392:353-358(1998).  
 CC FAMILY.  
 CC -1- SIMILARITY: TO PURINE/PYRIMIDINE PHOSPHORIBOSYL TRANSFERASE  
 CC FAMILY.  
 CC EMBL; AE000770; AAC07809.1; -  
 CC HSSP; P14193; IOKU.  
 DR INTERPRO; IPR002375; -  
 DR PROSITE; PS00103; PUR\_PYR\_PX\_TRANSFER; 1.  
 DR GLYCOSYLTRANSFERASE; Hypothetical protein; Transferase.  
 KW SEQUENCE 213 AA; 25205 MW; 883E9DED2F78309D CRC64;

Query Match 2.0%; Score 7; DB 2; Length 213;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 180 DLLFLSE 186  
 |||||  
 DB 9 DLLFLSE 15

RESULT 46  
 O41041 PRELIMINARY; PRT; 213 AA.  
 ID O41041

AC O41041;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE A559L PROTEIN.  
 GN A559L.  
 OS Paramacium bursaria chlorella virus 1 (PBCV-1).  
 OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Phycodnavirus.  
 OX NCBI\_TaxID=10506;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=98022962; PubMed=9356347;  
 RA Li Y., Lu Z., Sun L., Kopp S., Kutish G.F., Rock D.L., Van Etten J.L.;  
 RT "Analysis of 74 kb of DNA located at the right end of the 330-kb  
 RT chlorella virus PBCV-1 genome.";  
 RL Virology 237:360-377(1997).  
 DR EMBL; U42580; AAC6918.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 213 AA; 24034 MW; F23D014E68C07069 CRC64;

Query Match 2.0%; Score 7; DB 14; Length 213;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 VILALVA 12  
 |||||  
 DB 10 VILALVA 16

RESULT 47  
 O9FW32 PRELIMINARY; PRT; 220 AA.  
 AC O9FW32;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE HYPOTHEICAL 23.0 KDA PROTEIN.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae;  
 OC Oryza.  
 RN [1]  
 OX NCBI\_TaxID=4530;  
 RP SEQUENCE FROM N.A.  
 RA Hsing Y.-I.C., Chow T.-Y., Wu H.-P., Chao Y.-T., Liu S.-M.,  
 RA Chen C.-S., Shaw J.-F.;  
 RT "Oryza sativa PAC P0036D10 genomics sequence, complete sequence.";  
 RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AC073405; AAC03096.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 220 AA; 22958 MW; D8C72A0ADD4C780D CRC64;

Query Match 2.0%; Score 7; DB 10; Length 220;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 VILALVA 12  
 |||||  
 DB 7 VILALVA 13

RESULT 48  
 O9WY69 PRELIMINARY; PRT; 224 AA.  
 AC O9WY69;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE CONSERVED HYPOTHEICAL PROTEIN.  
 GN TMO226.  
 OS Thermotoga maritima.  
 OC Bacteria; Thermotogales; Thermotoga.

NCBI\_TaxID=2336;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MSB8 / DSM 3109;  
 MEDLINE=99287316; PubMed=10360571;  
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,  
 Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
 McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
 Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
 Heidelberg J., Sutton G.G., Fleischmann R.D., Eissen J.A., White O.,  
 Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
 genome sequence of *Thermotoga maritima*.";  
 RL Nature 399:323-329(1999);  
 DR EMBL: AE001707; AAC03518.1; -;  
 DR TIGR: TM0226; -;  
 SQ SEQUENCE 224 AA; 23912 MW; 95884FB5D24AFDF3 CRC64;

Query Match 2.0%; Score 7; DB 2; Length 224;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LLAIVAG 13  
 Db 61 LLAIVAG 67

RESULT 49  
 045081 PRELIMINARY; PRT; 225 AA.  
 AC 045081;  
 DT 01-JUN-1998 (TREMblrel. 06, Created)  
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)  
 DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)  
 DE CODED FOR BY C. ELEGANS CDNA YK127D8.5.  
 GN C17H12.10.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitoidea;  
 OC Rhabdilitidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M.,  
 Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,  
 Cratton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 Jones M., Kersey J., Kirsten J., Laister N., Latreille P.,  
 Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,  
 Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 Watson A., Weinstock L., Wilkinson-Spoat J., Wohlman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C.  
 elegans*.";  
 RL Nature 368:32-38(1994).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 Geisel C.;  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 Waterston R.;  
 RA Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF045642; AAC02588.1; -;  
 SQ SEQUENCE 225 AA; 25308 MW; 59D39C052F118C02 CRC64;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 238 KILVDAL 244  
 Db 73 KILVDAL 79

RESULT 50  
 0912N3 PRELIMINARY; PRT; 228 AA.  
 ID 0912N3;  
 AC 0912N3;  
 DT 01-MAR-2001 (TREMblrel. 16, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
 DE MOLYDENDIN TRANSPORT PROTEIN MODB.  
 GN MODB OR PA1862.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 Garber R.L., Colby L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an  
 opportunistic pathogen.";  
 RL Nature 406:959-964(2000).  
 DR EMBL: AE004612; AAC05251.1; -;  
 DR InterPro: IPR000515; -;  
 DR Pfam: PF00528; BPD\_transp; 1.  
 SQ SEQUENCE 228 AA; 24429 MW; 8223619B3B290E2C CRC64;

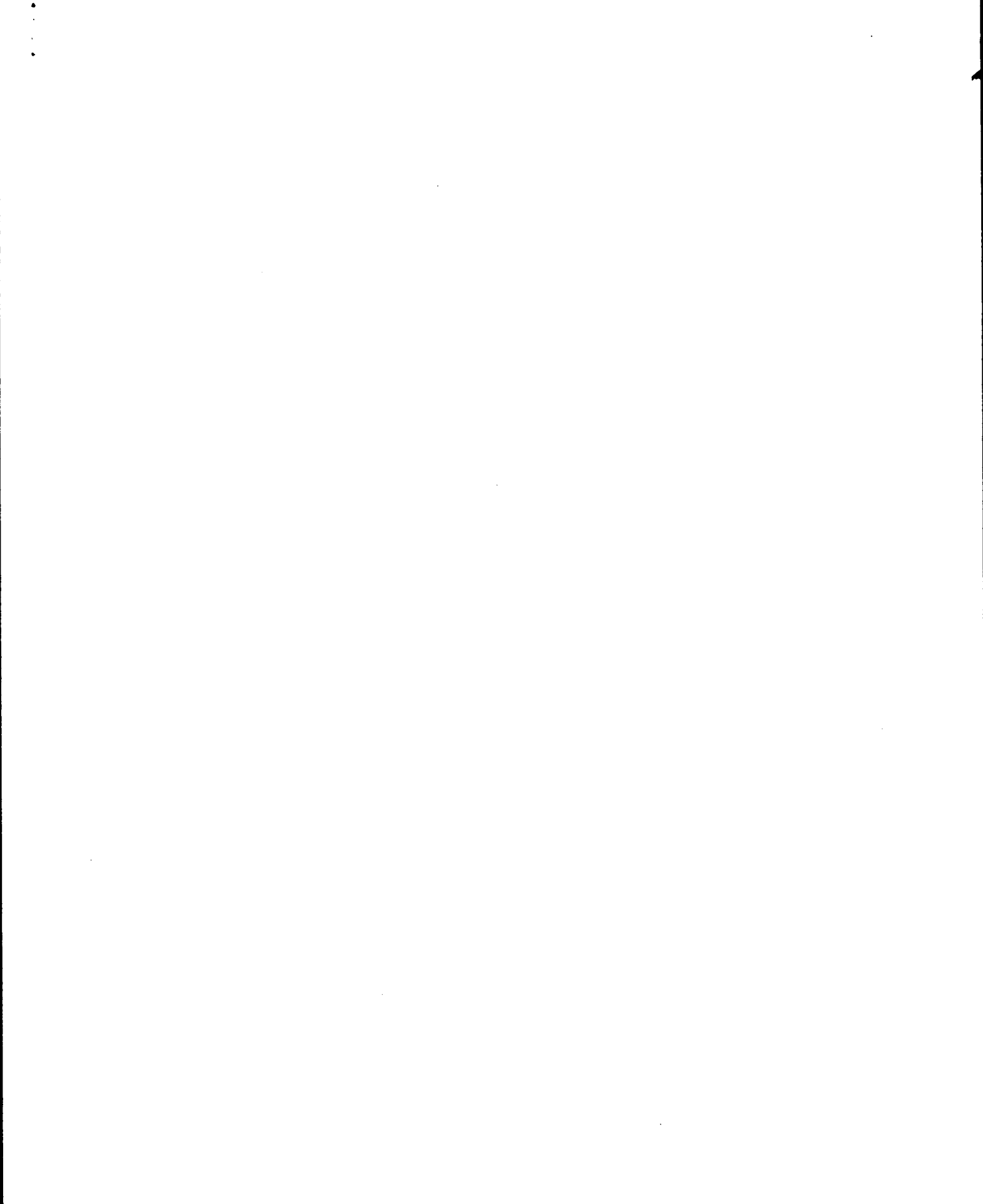
Query Match 2.0%; Score 7; DB 2; Length 228;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 FVLLAL 10  
 Db 211 FVLLAL 217

Search completed: August 17, 2001, 16:46:49  
 Job time: 151 sec

Query Match 2.0%; Score 7; DB 5; Length 225;























983 5 1.4 102 1 VEF\_HPV25 p36823 human papill  
984 5 1.4 102 1 Y66A\_HAEIN 086228 haemophilus  
985 5 1.4 103 1 CHLB\_EOAR p37845 equisetum s  
986 5 1.4 103 1 CHLB\_EOAR 022214 equisetum s  
987 5 1.4 103 1 G052\_MOUSE 061585 mus musculu  
988 5 1.4 103 1 RPOW\_ARCV 029033 archaeglob  
989 5 1.4 103 1 VEF\_HPV12 p36819 human papill  
990 5 1.4 103 1 VGH\_BPST1 p03648 bacterioph  
991 5 1.4 104 1 GAST\_CANFA p03353 canis faml  
992 5 1.4 104 1 PTLA\_STRMU p26426 streptococ  
993 5 1.4 104 1 RFA3\_SCHRO 092374 schizosacch  
994 5 1.4 104 1 R336\_CARBL p49181 caenorhabd  
995 5 1.4 104 1 RNC2\_ASPL 006652 aspergillus  
996 5 1.4 104 1 R310\_THEDC p28079 thermophilus  
997 5 1.4 104 1 Y04N\_BPT4 p39253 bacterioph  
998 5 1.4 104 1 Y04N\_BPT4 p39253 bacterioph  
999 5 1.4 104 1 Y04N\_BPT4 p39253 bacterioph  
1000 5 1.4 105 1 INS\_BOVIN p20090 thobacillu  
P01317 bos taurus

## ALIGNMENTS

RESULT 1  
VATN\_HUMAN STANDARD; PRT: 100 AA.  
ID VATN\_HUMAN  
AC 075787;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE VACUOLAR ATP SYNTHASE MEMBRANE SECTOR ASSOCIATED PROTEIN M8-9  
DE (V-ATPASE M8.9 SUBUNIT).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ludwig J., Schagger H., Boyd A., Apps D.K.;  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: VACUOLAR ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY  
CC OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS.  
CC -1- SUBUNIT: COMPOSED OF AT LEAST 10 SUBUNITS.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: Y17975; CAA76984.1;  
KW Hydroxylase; Hydrogen ion transport; Transmembrane.  
FT TRANSMEM 53  
FT TRANSSEM 53  
SQ SEQUENCE 100 AA: 11575 MW: A0705DB7B65F6DC CRC64;

Query Match 28.6%; Score 100; DB 1; Length 100;  
Best Local Similarity 100.0%; Pred. No. 3.8e-92;  
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 251 MSLYSGNAVVELVTVKSFDSILRTKRTILFAKAKKPPSPNLAAYKKNFYSVENMV 310  
DB 1 MSLYSGNAVVELVTVKSFDSILRTKRTILFAKAKKPPSPNLAAYKKNFYSVENMV 60  
OY 311 LWMIALALAVITTSYNTWMDPGYDSITTYRTNOKTRMD 350  
DB 61 LWMIALALAVITTSYNTWMDPGYDSITTYRTNOKTRMD 100

RESULT 2

VATN\_BOVIN STANDARD; PRT: 30 AA.  
ID VATN\_BOVIN  
AC P81134;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE VACUOLAR ATP SYNTHASE MEMBRANE SECTOR ASSOCIATED PROTEIN M8-9  
DE (V-ATPASE M8.9 SUBUNIT) (FRAGMENT).  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Adrenal medulla;  
RX MEDLINE=9822516; PUBMED=9556572;  
RA Ludwig J., Kersch S., Brandt U., Pfeiffer K., Getlawi F., Apps D.K.,  
RA Schagger H.;  
RT "Identification and characterization of a novel 9.2-kDa membrane  
RT sector-associated protein of vacuolar proton-ATPase from chromaffin  
RT granules.";  
RL J. Biol. Chem. 273:10939-10947(1998).  
CC  
CC -1- FUNCTION: VACUOLAR ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY  
CC OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS.  
CC -1- SUBUNIT: COMPOSED OF AT LEAST 10 SUBUNITS.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
KW Hydroxylase; Hydrogen ion transport; Transmembrane.  
FT NON\_TER 1  
FT TRANSEM 22  
FT NON\_TER 30  
FT NON\_TER 30  
SQ SEQUENCE 30 AA: 3585 MW: EA9FB26DCFA7D665 CRC64;

Query Match 3.1%; Score 11; DB 1; Length 30;  
Best Local Similarity 100.0%; Pred. No. 0.00051;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 293 YNLAKYKPFY 303  
DB 12 YNLAKYKPFY 22

RESULT 3  
TRMD\_MYCLE STANDARD; PRT: 238 AA.  
ID TRMD\_MYCLE  
AC O33017;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE TRNA (GUANINE-N1)-METHYLTRANSFERASE (EC 2.1.1.31) (MIG-  
DE METHYLTRANSFERASE) (TRNA [G37] METHYLTRANSFERASE).  
GN TRMD OR MLC8250.35.  
OS Mycobacterium leprae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1769;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Seeger K.J., Harris D., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
CC  
CC -1- FUNCTION: SPECIFICALLY METHYLATES GUANOSINE-37 IN VARIOUS TRNAS  
CC (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + TRNA - S-ADENOSYL-L-  
CC L-HOMOCYSTEINE + TRNA CONTAINING N(1)-METHYLGUANINE.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO THE RNA METHYLTRANSFERASE TRMD FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL: 297369; CAB10629.1; -  
 DR InterPro: IPR000531; -  
 DR InterPro: IPR002649; -  
 DR Pfam: PF01746; tRNA\_mig\_MT; 1.  
 DR Transferase: Methyltransferase; tRNA processing.  
 SO SEQUENCE 238 AA; 26229 MW; 784410B2BE7E320B CRC64;

Query Match 2.3%; Score 8; DB 1; Length 238;  
 Best Local Similarity 100.0%; Pred. No. 3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 10 LVAGVLGN 17  
 |||||  
 Db 155 LVAGVLGN 162

RESULT 4  
 GON3\_PORNO STANDARD; PRT: 89 AA.  
 AC P51922;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE GONADOLIBERIN III PRECURSOR (GONADOTROPIN-RELEASING HORMONE III)  
 DE (GNRH-III) (LH-RH III) (LULIBERIN III).  
 OS Porichthys notatus (Platfish midshipman).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Paracanthopterygii; Batrachoididae; Porichthys.  
 NCBI\_TaxID=45384;  
 [1]  
 RN RP  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95385993; PubMed=7657161;  
 RA Grober M.S., Myers T.R., Marchette M.A., Bass A.H., Myers D.A.;  
 RT "Structure, localization, and molecular phylogeny of a GNRH cDNA from  
 RT a paracanthopterygian fish, the platfish midshipman (Porichthys  
 RT notatus).";  
 RL Gen. Comp. Endocrinol. 99:85-99(1995).  
 CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS.  
 CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: U41669; AAC59754.1; -  
 DR EMBL: S79620; AAB35188.1; -  
 DR InterPro: IPR002012; -  
 DR Pfam: PF00446; GNRH; 1.  
 DR PROSITE: PS00473; GNRH; 1.  
 KW Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;  
 KW Signal.

FT SIGNAL 1 23 BY SIMILARITY.  
 FT CHAIN 24 89 PROGNADOLIBERIN III.  
 FT PEPTIDE 24 33 GONADOLIBERIN III.  
 FT PEPTIDE 37 89 GNRH-ASSOCIATED PEPTIDE III (POTENTIAL).  
 FT MOD\_RES 24 24 PYROLIDONE CARBOXYLIC ACID  
 FT MOD\_RES (BY SIMILARITY).  
 FT MOD\_RES AMIDATION (G-34 PROVIDE AMIDE GROUP).  
 SO SEQUENCE 89 AA; 10118 MW; AED532789B9F1475 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 89;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 VVLLALV 11  
 |||||  
 Db 11 VVLLALV 17

RESULT 5  
 TX1A\_AGEAP STANDARD; PRT: 112 AA.  
 ID TX1A\_AGEAP  
 AC P15969;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE OMEGA-AGATOXIN IA PRECURSOR (OMEGA-AGA-IA).  
 OS Agelenopsis aperta (Funnel-web spider).  
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;  
 OC Araneomorphae; Entelegynae; Agelenidae; Agelenopsis.  
 NCBI\_TaxID=6908;  
 [1]  
 RN RP  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE-Venom;  
 RX MEDLINE=93015969; PubMed=1383207;  
 RA Santos A.D., Imperial J.S., Chaudhary T., Beavis R.C., Chait B.T.,  
 RA Hunsberger J.P., Olivera B.M., Adams M.E., Hilliard D.R.;  
 RT "Heterodimeric structure of the spider toxin omega-agatoxin IA  
 RT revealed by precursor analysis and mass spectrometry.";  
 RL J. Biol. Chem. 267:20701-20705(1992).  
 [2]  
 RN RP  
 RP SEQUENCE OF 37-102.  
 RC TISSUE-Venom;  
 RX MEDLINE=90110147; PubMed=2295621;  
 RA Adams M.E., Bindokas V.P., Hasegawa L., Venema V.J.;  
 RT "Omega-agatoxins: novel calcium channel antagonists of two subtypes  
 RT from funnel web spider (Agelenopsis aperta) venom.";  
 RL J. Biol. Chem. 265:861-867(1990).  
 CC -1- FUNCTION: OMEGA-AGATOXIN ARE ANTAGONIST OF VOLTAGE-SENSITIVE  
 CC CALCIUM CHANNELS. THEY BLOCK INSECT NEUROMUSCULAR TRANSMISSION  
 CC PRESYNAPTICALLY.  
 CC -1- SUBUNIT: HETERODIMER OF TWO SUBUNITS, A MAJOR CHAIN AND A MINOR  
 CC CHAIN, LINKED BY A DISULFIDE BOND.  
 CC -1- PTM: THE TWO SUBUNITS ARE PROCESSED FROM THE PRECURSOR PEPTIDE  
 CC BY THE EXCISION OF AN INTERNAL PEPTIDE OF 7 RESIDUES.  
 CC -1- MASS SPECTROMETRY: MW=7791.4; METHOD=PLASMA DESORPTION;  
 CC RANGE=37-102;110-112.  
 CC -1- SIMILARITY: BELONGS TO THE MU-AGATOXIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M95540; -; NOT ANNOTATED\_CDS.  
 DR PIR: A34923; A34923.  
 DR Venom; Neurotoxin; Calcium channel inhibitor; Presynaptic neurotoxin;  
 KW Venom; Neurotoxin; Calcium channel inhibitor; Presynaptic neurotoxin;  
 KW Signal.

FT SIGNAL 1 19 POTENTIAL.  
 FT PROPEP 20 36  
 FT CHAIN 37 102 OMEGA-AGATOXIN-IA MAJOR CHAIN.  
 FT PROPEP 103 109 GLU-RICH.  
 FT CHAIN 110 112 OMEGA-AGATOXIN-IA MINOR CHAIN.  
 FT CHAIN 101 101 D -> N (IN REF. 2).  
 FT CONFLICT 101 101  
 SO SEQUENCE 112 AA; 12808 MW; ACC0ED76E49DF23 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 112;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 39 ALPGSV 45

## RESULT 6

YIJD\_ECOLI STANDARD; PRT; 119 AA.

AC P27308;

DT 01-APR-1992 (Rel. 23, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)

DE 01-NOV-1997 (Rel. 35, Last annotation update)

DE HYPOTHETICAL 13.0 KDA PROTEIN IN UDHA-TRMA INTERGENIC REGION (ORFB).

GN YIJD.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI\_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-K12 / MG1655;

RX MEDLINE=94089392; PubMed=8265357;

RA Blatter F.R., Burland V.D., Plunkett G. III, Sofia H.J.,

RA Daniels D.L.,

RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the

RT region from 89.2 to 92.8 minutes."

RI Nucleic Acids Res. 21:5408-5417(1993).

RL [2]

RN SEQUENCE FROM N.A.

RC STRAIN-K12;

RX MEDLINE=93077482; PubMed=1447162;

RA Gustafsson C., Warne S.R.;

RA "Physical map of the oxyR-trmA region (minute 89.3) of the

RT Escherichia coli chromosome."

RT J. Bacteriol. 174:7878-7879(1992).

RL [1]

RN SEQUENCE FROM N.A.

RC EMBL; U00006; AAC43070.1; -

DR EMBL; AE000470; AAC76946.1; -

DR EMBL; X66026; CAA46824.1; -

DR PIR; S21565; S21565.

DR Ecogene; BG11395; yljD.

RW Hypothetical protein.

FT CONFLICT 13 13

FT CONFLICT 94 94

FT CONFLICT 100 110

SQ SEQUENCE 119 AA; 13023 MW; 29177570EAF20F3B CRC64;

Query Match 2.0%; Score 7; DB 1; Length 119;

Best Local Similarity 100.0%; Pred. No. 16;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 LLAIVAG 13

Db 12 LLAIVAG 18

## RESULT 7

Y074\_MYCE STANDARD; PRT; 137 AA.

AC P47320;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DE HYPOTHETICAL PROTEIN MG074.

GN MG074.

OS Mycoplasma genitalium.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
OC Mycoplasmataceae; Mycoplasma.

OX NCBI\_TaxID=2097;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ATCC 33530 / G-37;

RX MEDLINE=96026346; PubMed=7569993;

RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,

RA Fleischmann R.D., Fulton C.J., Kerlavage A.R., Sutton G., Kelley J.M.,

RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,

RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,

RA Tomb J.-F., Dougherty B.A., Bort K.F., Hu P.-C., Lucier T.S.,

RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;

RT "The minimal gene complement of Mycoplasma genitalium."

RL Science 270:397-403(1995).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U39687; AAC71292.1; -

DR TIGR; MG074; -

RW Hypothetical protein.

SQ SEQUENCE 137 AA; 16616 MW; 25D8FDD2803126E7 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 137;

Best Local Similarity 100.0%; Pred. No. 18;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 181 LFLFSL 187

Db 14 LFLFSL 20

## RESULT 8

YB10\_METH

AC 027182;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DE HYPOTHETICAL PROTEIN MTH110.

GN MTH110.

OS Methanobacterium thermoautotrophicum.

OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;

OC Methanothermobacter.

OX NCBI\_TaxID=145262;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-DELTA H;

RX MEDLINE=98037514; PubMed=9371463;

RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., Dubois J.,

RA Aldridge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,

RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,

RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,

RA Jivani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,

RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,

RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;

RT "Complete genome sequence of Methanobacterium thermoautotrophicum

RT deltaH: functional analysis and comparative genomics."

RL J. Bacteriol. 179:7135-7155(1997).

CC -1- SIMILARITY: TO M.JANNASCHII MJ1244 AND MJ1245.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entries requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL: AE000881; AAB85599.1; -  
CC Hypothetical protein.  
KM SEQUENCE 151 AA; 17176 MW; 874FB20F033D081B CRC64;

Query Match 2.0%; Score 7; DB 1; Length 151;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 51 GFSYKED 57  
DB 87 GFSYKED 93

## RESULT 9

FLAP\_MOUSE STANDARD; PRT; 153 AA.

AC P30355;  
DT 01-APR-1993 (rel. 25, Created)  
DT 01-APR-1993 (rel. 25, Last sequence update)  
DT 01-APR-2000 (rel. 40, Last annotation update)  
DE 5-LIPOXYGENASE ACTIVATING PROTEIN (FLAP) (MK-886-BINDING PROTEIN)  
DE (FRAGMENT).  
GN ALOX5AP OR FLAP.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Vickers P.J., O'Neill G.P., Mancini J.A., Charleson S.,  
RA Abramovitz M.;  
RL Submitted (XXX-1992) to the EMBL/Genbank/DBJ databases.  
CC -1- FUNCTION: SEEMS TO BE REQUIRED FOR THE ACTIVATION OF 5-LO (5-  
LIPPOXYGENASE). FLAP COULD PLAY AN ESSENTIAL ROLE IN THE TRANSFER  
OF ARACHIDONIC ACID TO 5-LO. FLAP BINDS TO MK-886, A COMPOUND  
THAT BLOCKS THE BIOSYNTHESIS OF LEUKOTRIENES.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO THE MAPEG FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: M96554; AAA37632.1; -  
DR MGD: MGI:107505; Alox5ap.  
DR InterPro: IPR001129; -  
DR Pfam: PF01124; MAPEG; 1.  
DR PROSITE: PS01297; FLAP\_GST\_LTCAS; 1.  
FT TRANSMEMBRANE; Leukotriene biosynthesis.  
FT TRANSMEM 6 26  
FT TRANSMEM 67 87 POTENTIAL.  
FT TRANSMEM 118 138 POTENTIAL.  
FT NON\_TER 153 153  
SQ SEQUENCE 153 AA; 17337 MW; E611ABF1A3604AF1 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 151;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 VVLLALV 11  
DB 9 VVLLALV 15

RESULT 10  
ID PHCA\_SYNP6 STANDARD; PRT; 162 AA.

AC P00308;  
DT 21-JUL-1986 (rel. 01, Created)  
DT 21-JUL-1986 (rel. 01, Last sequence update)  
DT 01-OCT-2000 (rel. 40, Last annotation update)  
DE C-PHYCOCYANIN ALPHA CHAIN.  
GN CPCH.  
OS Synechococcus sp. (strain PCC 6301) (Anacystis nidulans).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
OX NCBI\_TaxID=1139;  
RN [1]  
RP SEQUENCE.

RA Walsh R.G., Wingfield P., Glazer A.N., Delange R.J.;  
RT "Amino acid sequence of C-phycoerythrin, alpha-subunit, from  
RT Synechococcus 6301 (Anacystis nidulans).";  
RL Fed. Proc. 39:1998-1998(1980).  
CC -1- FUNCTION: LIGHT-HARVESTING PHOTOSYNTHETIC BILE PIGMENT-PROTEIN  
CC FROM THE PHYCOBILIPROTEIN COMPLEX.  
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.  
CC -1- PTM: CONTAINS ONE COVALENTLY LINKED BILIN CHROMOPHORE.  
DR PIR: A00317; CFYCA.  
DR HSSP: P07122; ICPC.  
DR InterPro: IPR001659; -  
DR Pfam: PF00502; Phycobilisome; 1.  
KW Phycobilisome; Electron transport; Photosynthesis; Bile pigment.  
FT BINDING 84 84 PHYCOCYANOBILIN CHROMOPHORE.  
SQ SEQUENCE 162 AA; 17158 MW; 5792441D5C676C74 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 162;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 216 LAGIDEI 222  
DB 112 LAGIDEI 118

## RESULT 11

PHCA\_SYNP7 STANDARD; PRT; 162 AA.

AC P13530; Q04120;  
DT 01-JAN-1990 (rel. 13, Created)  
DT 01-JAN-1990 (rel. 13, Last sequence update)  
DT 15-JUL-1998 (rel. 36, Last annotation update)  
DE C-PHYCOCYANIN ALPHA CHAIN.  
GN CPCH OR CPCHL.  
OS Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
OX NCBI\_TaxID=1140;  
RN [1]  
RP SEQUENCE FROM N.A.

RA MEDLINE=87248092; PubMed=3036657;  
RA Lau R.H., Alvarado-urbina G., Lau P.C.K.;  
RT "Phycocyanin alpha-subunit gene of Anacystis nidulans R2: cloning,  
RT nucleotide sequencing and expression in Escherichia coli.";  
RL Gene 52:21-29(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88257006; PubMed=2454910;  
RX Kalla R.S., Lind L.K., Lidholm J., Gustafsson P.;  
RT "Transcriptional organization of the phycocyanin subunit gene  
RT clusters of the cyanobacterium Anacystis nidulans UTEX 625.";  
RL J. Bacteriol. 170:2961-2970(1988).  
RN [3]  
RP SEQUENCE OF 1-11 FROM N.A.

RX MEDLINE=87174767; PubMed=3104880;  
RX Lau P.C.K., Condie A., Alvarado-urbina G., Lau R.H.;  
RT "Nucleotide sequence of phycocyanin beta-subunit gene of  
RT cyanobacterium Anacystis nidulans strain R2.";  
RL Nucleic Acids Res. 15:2394-2394(1987).

```

CC -1- FUNCTION: LIGHT-HARVESTING PHOTOSYNTHETIC BILE PIGMENT-PROTEIN
CC FROM THE PHYCOBILIPROTEIN COMPLEX.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
CC -1- PTM: CONTAINS ONE COVALENTLY LINKED BILIN CHROMOPHORE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M16325; AAA22051.1; -
DR EMBL: M94218; AAA64532.1; -
DR EMBL: M94218; CAB28568.1; -
DR EMBL: X04916; CAA28586.1; -
DR PIR: A29015; A29015.
DR HSSP: P07122; IOPC.
DR InterPro: IPR001659; -
DR Pfam: PF00502; Phycobilisome; 1.
DR Phycobilisome; Electron transport; Photosynthesis; Bile pigment.
DR INIT_MET 0 0
DR BINDING 84 84 PHYCOCYANOBILIN CHROMOPHORE.
DR SEQUENCE 162 AA; 17157 MW; 6CE6C341B1B3CFE CRC64;
SQ
Query Match 2.0%; Score 7; DB 1; Length 162;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 216 LAGLDEI 222
DB 112 LAGLDEI 118
RESULT 12
CLD5_RAT
ID CLD5_RAT STANDARD; PRT; 206 AA.
AC 09JRD6;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CLAUDIN-5 (FRAGMENT).
GN CLDN5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID-10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Kirsch T., Wellner M., Haller H., Lippold A.;
RT "Cloning of rat claudin-5."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF241260; AAF3425.1; -
DR InterPro: IPR001832; -
DR InterPro: IPR002059; -
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN.1.
KW Tight junction; Transmembrane.
NON_TER 1 1

```

```

FT TRANSMEM <1 19 POTENTIAL.
FT TRANSMEM 73 93 POTENTIAL.
FT TRANSMEM 115 135 POTENTIAL.
FT TRANSMEM 152 172 POTENTIAL.
FT NON_TER 206 206
SQ SEQUENCE 206 AA; 21810 MW; F640B544E6E82F2 CRC64;
Query Match 2.0%; Score 7; DB 1; Length 206;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 VLLALVA 12
DB 79 VLLALVA 85
RESULT 13
CLD5_MOUSE
ID CLD5_MOUSE STANDARD; PRT; 218 AA.
AC 054942; 088789;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-5 (BRAIN ENDOTHELIAL CELL CLONE 1 PROTEIN) (LUNG-SPECIFIC
DE MEMBRANE PROTEIN).
GN CLDN5 OR BECL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID-10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Tissue-Brain;
RC MEDLINE-98181565; PubMed-9520948;
RA Chen Z., Zandonati M., Jakubowski D., Fox H.S.;
RT "Brain capillary endothelial cells express MBEC1, a protein that is
RT related to the Clostridium perfringens enterotoxin receptors."
RL Lab. Invest. 78:353-363(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Kim K.K., Baek D.H., Kwon B.S.;
RT "A mouse lung-specific membrane protein."
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA MEDLINE-99110921; PubMed-9892664;
RA Morita K., Furuse M., Fujimoto K., Tsukita S.;
RT "Claudin multigene family encoding four-transmembrane domain protein
RT components of tight junction strands."
RL Proc. Natl. Acad. Sci. U.S.A. 96:511-516(1999).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF035814; AAB96653.1; -
DR EMBL: U82758; AAC27545.1; -
DR EMBL: AF087823; AAD09758.1; -
DR MGD: MGI:1276112; Cldn5.
DR InterPro: IPR000729; -
DR InterPro: IPR001832; -
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN.1.
KW Tight junction; Transmembrane.
NON_TER 1 1

```

FT TRANSMEM 8 28 POTENTIAL.  
 FT TRANSMEM 82 102 POTENTIAL.  
 FT TRANSMEM 124 144 POTENTIAL.  
 FT TRANSMEM 161 181 POTENTIAL.  
 FT CONFLICT 58 58 S -> N (IN REF. 1).  
 SO SEQUENCE 218 AA; 23054 MW; 8708F69AE282DE87 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 218;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 VLLALVA 12  
 |||||  
 DB 88 VLLALVA 94

RESULT 14  
 RNPL\_ECOLI

ID RNPL\_ECOLI STANDARD; PRT; 238 AA.

AC P03842;

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE RIBONUCLEASE PH (EC 2.7.7.56) (RNASE PH) (TRNA

DE NUCLEOTIDYLTRANSFERASE).

GN RPH.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia

OX NCBI\_TaxID=562;

RN (1)

RP SEQUENCE FROM N.A.

RC MEDLINE=85003588; PubMed=6207018;

RA Poulsen P., Bonekamp F., Jensen K.F.;

RT "Structure of the Escherichia coli pyre operon and control of pyre

RT expression by a uvp modulated interictronic attenuation.";

RL EMBO J. 3:1783-1790(1984).

RN (2)

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;

RA MEDLINE=93315143; PubMed=7686882;

RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli

RT genome: organizational symmetry around the origin of replication.";

RL Genomics 16:551-561(1993).

RN (3)

RP IMPORTANCE AT HIGH TEMPERATURE.

RC MEDLINE=92021791; PubMed=1925027;

RA Poulsen P., Jensen K.F.;

RT "Three genes preceding pyre on the Escherichia coli chromosome are

RT essential for survival and normal cell morphology in stationary

RT culture and at high temperature.";

RL Res. Microbiol. 147:283-288(1991).

RN (4)

RP FUNCTION, AND PARTIAL SEQUENCE OF 1-21.

RC MEDLINE=91358344; PubMed=1885537;

RA Ost K.A., Deutscher M.P.;

RT "Escherichia coli orfE (upstream of pyre) encodes RNase PH.";

RL J. Bacteriol. 173:5589-5591(1991).

RN (5)

RP CHARACTERIZATION.

RC MEDLINE=92381028; PubMed=1512252;

RA Jensen K.F., Andersen J.F., Poulsen P.;

RT "Overexpression and rapid purification of the orfE/rph gene product,

RT RNase PH of Escherichia coli.";

RL J. Biol. Chem. 267:17147-17152(1992).

RN (6)

RP CHARACTERIZATION.

RC MEDLINE=92381029; PubMed=1512253;

RA Kelly K.O., Deutscher M.P.;

RT "Characterization of Escherichia coli RNase PH.";

RL J. Biol. Chem. 267:17153-17158(1992).

CC -I- FUNCTION: RNASE PH IS A PHOSPHORYLYTIC EXORIBONUCLEASE THAT  
 CC REMOVES NUCLEOTIDE RESIDUES FOLLOWING THE -CCA TERMINUS OF TRNA  
 CC AND ADDS NUCLEOTIDES TO THE ENDS OF RNA MOLECULES BY USING  
 CC NUCLEOSIDE DIPHOSPHATES AS SUBSTRATES. ALSO ACTS TO REGULATE THE  
 CC ATTENUATION OF PYRE.  
 CC -I- CATALYTIC ACTIVITY: TRNA(N+1) + ORTHOPHOSPHATE = TRNA(N) +  
 CC A NUCLEOSIDE DIPHOSPHATE  
 CC -I- SIMILARITY: BELONGS TO THE RNASE PH FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X00781; CAA25357.1; -;  
 DR EMBL: V01578; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL: L10328; AAA61996.1; -;  
 DR EMBL: AE000441; AAC76667.1; -;  
 DR PIR: A04470; OOECP.  
 DR Ecogene: EG10863; rph.  
 DR InterPro: IPR001247; -;  
 DR InterPro: IPR002381; -;  
 DR Pfam: PF01138; RNase\_PH; 1.  
 DR PROSITE: PS01277; RIBONUCLEASE\_PH; 1.  
 KW Transferase; Nucleotidyltransferase; RNA processing.  
 FT CONFLICT 123 124 GG -> AW (IN REF. 1).  
 FT CONFLICT 224 238 GIESIVATOKALAN -> ESNPL (IN REF. 2).  
 SO SEQUENCE 238 AA; 25352 MW; 7299C5C8F2364D34 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 238;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 240 LVDALOK 246  
 |||||  
 DB 137 LVDALOK 143

RESULT 15  
 OPSD\_COTBO

ID OPSD\_COTBO STANDARD; PRT; 289 AA.

AC 042307;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE RHODOPSIN (FRAGMENT).

GN RHO.

OS Coltinella boulegeri.

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Acanthopterygii; Percomorpha; Scorpaeniformes;

OC Cottoidei; Aduosocottidae; Cottinella.

OX NCBI\_TaxID=61639;

RN (1)

RP SEQUENCE FROM N.A.

RC MEDLINE=98086781; PubMed=9417898;

RA Hunt D.M., Fitzgibbon J., Slobodyanyuk S.J., Bowmaker J.K.;

RT "Molecular evolution of the cottoid fish endemic to Lake Baikal

RT deduced from nuclear DNA evidence.";

RL Mol. Phylogenet. Evol. 8:415-422(1997).

CC -I- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT

CC MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY

CC LINKED TO CIS-RETINAL.

CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

CC -I- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES

CC VISION IN DIM LIGHT.

CC -I- PTM: SOME OR ALL OF THE CARBOXY-TERMINAL SER OR THR RESIDUES MAY

CC BE PHOSPHORYLATED (BY SIMILARITY).

```

CC -1 SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC OPSIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U97273; AAB6127.1; .
DR GCRDB: GCR1260; .
DR InterPro: IPR000276; .
DR InterPro: IPR001760; .
DR Pfam: PF00001; 7tm_1; 1.
DR PROSITE: PS00237; G-PROTEIN_RECEP_F1_1; 1.
DR PROSITE: PS50262; G-PROTEIN_RECEP_F1_2; 1.
DR PROSITE: PS00238; OPSIN; 1.
KW Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
KW Phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor.
FT NON_TER 1 1
FT DOMAIN 1 7 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 8 32 1 (POTENTIAL).
FT DOMAIN 33 44 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 45 69 2 (POTENTIAL).
FT DOMAIN 70 84 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 85 104 3 (POTENTIAL).
FT DOMAIN 105 123 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 124 147 4 (POTENTIAL).
FT DOMAIN 148 173 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 174 201 5 (POTENTIAL).
FT DOMAIN 202 223 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 224 247 6 (POTENTIAL).
FT DOMAIN 248 255 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 256 280 7 (POTENTIAL).
FT DOMAIN 281 >289 CYTOPLASMIC (POTENTIAL).
FT DISULFID 81 158 BY SIMILARITY.
FT BINDING 267 267 RETINAL CHROMOPHORE (BY SIMILARITY).
FT CARBOHD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 289 289
SQ SEQUENCE 289 AA; 32753 MW; FDFEHEECF10445D4 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 289;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 LAVGNLF 69
| | | | |
Db 50 LAVGNLF 56

RESULT 16
YDCT_ECOLI STANDARD; PRT; 337 AA.
ID YDCT_ECOLI
AC P77795;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YDCT.
GN YDCT.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=9742617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

```

```

RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
RA Kitagawa M., Makino K., Mori T., Mizobuchi K., Mori H., Mori T.,
RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
RA Oshima T., Saito N., Sampei G., Seki Y., Sivasubram S.,
RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
CC -1 FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT
CC SYSTEM YDCTUV. PROBABLY RESPONSIBLE FOR ENERGY COUPLING TO THE
CC TRANSPORT SYSTEM.
CC -1 SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000241; AAC74523.1; .
DR EMBL: D90783; BAA15070.1; .
DR EMBL: D90784; BAA15076.1; .
DR Ecogene: EG13763; YDCT.
DR InterPro: IPR001617; .
DR Pfam: PF00005; ABC_tran; 1.
DR PROSITE: PS00211; ABC-TRANSPORTER; 1.
KW Hypothetical protein; ATP-binding; Transport.
KW NP_BIND 37 44
SQ SEQUENCE 337 AA; 37040 MW; FAA179089645F14 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 337;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 MYKGVNK 85
| | | | |
Db 102 MYKGVNK 108

RESULT 17
CTGF_HUMAN STANDARD; PRT; 349 AA.
ID CTGF_HUMAN
AC P29279;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE CONNECTIVE TISSUE GROWTH FACTOR PRECURSOR.
GN CTGF.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=umbilical vein endothelial cells;
RX MEDLINE=91373452; PubMed=1654338;
RA Bradham D.M., Igarashi A., Potter R.L., Grotendorst G.R.;
RA "Connective tissue growth factor: a cysteine-rich mitogen secreted by
RA human vascular endothelial cells is related to the SRC-induced
RA immediate early gene product CEF-10.";

```



RL J. Cell Biol. 114:1285-1294(1991).  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-umbilical vein endothelial cells;  
 RA MEDLINE-9318714; PubMed-1293144;  
 RA Igarashi A., Bradham D.M., Okochi H., Grotendorst G.R.;  
 RT "Connective tissue growth factor.";  
 RN J. Dermatol. 19:642-643(1992).  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE-9720746; PubMed-9054739;  
 RA Oemar B.S., Werner A., Garnier J.M., Do D.D., Godoy N., Nauck M.,  
 Marz W., Rupp J., Pech M., Luescher T.F.;  
 RT "Human connective tissue growth factor is expressed in advanced  
 RT atherosclerotic lesions.";  
 RL Circulation 95:831-839(1997).  
 CC -1- FUNCTION: MAJOR CONNECTIVE TISSUE MITOCHONDRIANT SECRETED BY  
 CC HUMAN VASCULAR ENDOTHELIAL CELLS. THIS IMMEDIATE-EARLY PROTEIN  
 CC MAY BIND ONE OF THE PDGF CELL SURFACE RECEPTORS.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A  
 CC SHORT FORM; SEEM TO BE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING  
 CC PROTEIN FAMILY. CEF-10/CYR61/CTRG/ETSP-12/NOV PROTEIN SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: M92934; AAA91279.1; -  
 DR EMBL: X78947; CA55544.1; -  
 DR PIR: A40551; A40551.  
 DR PIR: S44205; S44205.  
 DR MIM: 121009; -  
 DR InterPro: IPR0000359; -  
 DR InterPro: IPR0000867; -  
 DR InterPro: IPR000884; -  
 DR InterPro: IPR001007; -  
 DR Pfam: PF00007; Cys\_knot; 1.  
 DR Pfam: PF00219; IGFBP; 1.  
 DR Pfam: PF00090; tsp\_1; 1.  
 DR Pfam: PF00093; wvc; 1.  
 DR PROSITE: PS00222; IGF\_BINDING; 1.  
 DR PROSITE: PS01185; CTCK\_1; 1.  
 DR PROSITE: PS01225; CTCK\_2; 1.  
 DR PROSITE: PS01208; WVC; 1.  
 KW Growth factor binding; signal; Alternative splicing.  
 FT SIGNAL 1 26 POTENTIAL.  
 FT CHAIN 1 26 CONNECTIVE TISSUE GROWTH FACTOR.  
 FT DOMAIN 101 167 WVC.  
 FT DISUFID 256 330 CTCK.  
 FT DISUFID 273 307 BY SIMILARITY.  
 FT DISUFID 284 323 BY SIMILARITY.  
 FT DISUFID 287 325 BY SIMILARITY.  
 FT DISUFID 292 329 BY SIMILARITY.  
 FT CARBOHYD 28 28 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 225 225 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPIC 172 198 MISSING (IN SHORT ISOFORM).  
 SQ SEQUENCE 349 AA; 38069 MW; 0BCF6470B357EA95 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 349;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 4 FVVLAL 10

Db 13 FVVLAL 19  
 |||||  
 RESULT 18  
 AMPN\_ACETU STANDARD; PRT; 355 AA.  
 ID AMPN\_ACETU  
 AC 010736;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE AMINOPEPTIDASE N (EC 3.4.11.2) (ALPHA-AMINOACYLPEPTIDE HYDROLASE)  
 DE (FRAGMENT).  
 GN PEPN.  
 OS Acetobacter turbidans.  
 OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;  
 OC Acetobacter.  
 OC NCBI\_TaxID=45671;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 9325;  
 RA Weber M.;  
 RT "Cloning and DNA sequence analysis of the 5' end of aminopeptidase N  
 RT (pepn) from Acetobacter turbidans ATCC9325.";  
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: AMINOPEPTIDASE N IS INVOLVED IN THE DEGRADATION OF  
 CC INTRACELLULAR PEPTIDES GENERATED BY PROTEIN BREAKDOWN DURING  
 CC NORMAL GROWTH AS WELL AS IN RESPONSE TO NUTRIENT STARVATION (BY  
 CC SIMILARITY).  
 CC -1- COFACTOR: BINDS ONE ZINC ION.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M1 (ZINC METALLOPROTEASE);  
 CC ALSO KNOWN AS THE PEPN SUBFAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: X94692; CAA64355.1; -  
 DR MEROPS: M01.005; -  
 DR InterPro: IPR00130; -  
 DR InterPro: IPR001930; -  
 DR Pfam: PF01433; Peptidase\_M1; 1.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; 1.  
 KW Hydrolyase; Metalloprotease; Aminopeptidase; Zinc.  
 FT METAL 326 326 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 327 327 BY SIMILARITY.  
 FT METAL 330 330 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 349 349 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 355 355 ZINC (CATALYTIC) (BY SIMILARITY).  
 SQ SEQUENCE 355 AA; 38820 MW; 8F214BDB2B368232 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 355;  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 7 LIALVAG 13  
 |||||  
 Db 219 LIALVAG 225  
 |||||  
 RESULT 19  
 HITC\_HAEIN STANDARD; PRT; 356 AA.  
 ID HITC\_HAEIN  
 AC P44513; Q53441;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)

```

DE IRON(III)-TRANSPORT ATP-BINDING PROTEIN HITC.
OS HTIC OR H10099.
OC Hemophilus influenzae.
CC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
CC Hemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,
RA Fine L.D., Fitchman J.L., Fuhmann J.L., Geophagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd. ";
RL Science 269:496-512(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WTIH TN106;
RX MEDLINE=95012644; PubMed=7927717;
RA Sanders J.D., Cope L.D., Hansen E.J.;
RT Identification of a locus involved in the utilization of iron by
RT Haemophilus influenzae.
RL Infect. Immun. 62:4515-4525(1994).
CC -1- FUNCTION: NICKLEOTIDE-BINDING. PROTEIN COMPONENT OF THE PERIPLASMIC
CC BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM FOR FE(III).
CC -1- SUBCELLULAR LOCATION: INNER MEMBRANE-ASSOCIATED (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U32695; AAC21775.1;
DR EMBL: S72674; AAB32112.1;
DR HSSP: P13569; INBD.
DR TIGR: H10099;
DR InterPro: IPR001617;
DR Pfam: PF00005; ABC_tran; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
KW Iron transport; Inner membrane; Transport; ATP-binding.
FT NP_BIND 44 51 ATP (POTENTIAL).
FT VARIANT 45 45 S -> A (IN STRAIN TN106).
FT VARIANT 97 97 V -> I (IN STRAIN TN106).
FT VARIANT 209 209 S -> A (IN STRAIN TN106).
FT VARIANT 350 350 R -> K (IN STRAIN TN106).
FT VARIANT 356 356 S -> A (IN STRAIN TN106).
SQ SEQUENCE 356 AA; 40360 MW; 78F862F359760813 CRC64;

```

```

AC 018347;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHELICAL 41.5 KDA PROTEIN C31H2.4 IN CHROMOSOME X.
GN C31H2.4.
OS Caenorhabditis elegans.
CC Caenorhabditis elegans.
CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditoidea;
CC Rhabditidae; Pelodirinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Geisel C., Gattung S.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE 4HPPD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U41748; AAB3338.1;
DR WormPeP: C31H2.4; CE04132.
DR InterPro: IPR002887;
DR Pfam: PF01626; 4HPPD_C; 1.
KW Hypothetical protein.
SQ SEQUENCE 364 AA; 41456 MW; 599310F028F28E18 CRC64;

```

Query Match 2.0%; Score 7; DB 1; Length 364;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 321 VIITSYN 327
|111111|
Db 53 VIITSYN 59

```

RESULT 21

```

CYCL_HUMAN
ID CYCL_HUMAN STANDARD; PRT; 377 AA.
AC Q14094;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CYCLIN I.
GN CCNT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Forebrain cortex;
RX MEDLINE=96086776; PubMed=7493655;
RA Nakamura T., Sanokawa R., Sasaki Y.F., Ayusawa D., Oishi M., Mori N.;
RT "Cyclin I: a new cyclin encoded by a gene isolated from human brain.";
RL Exp. Cell Res. 221:534-542(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Chen J.H., Luo W.O., Zhou Y., Zhou H.J., Huang X.W., Yuan J.G.,
RA Jiang B.Q.;
RT "Isolating a new cDNA coding for human cyclin protein.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-178 FROM N.A.
RC TISSUE-Testis;
RX MEDLINE=98381026; PubMed=9705831;
RA Zhu X., Naz R.K.;

```

RT "Expression of a novel isoform of cyclin I in human testis.";  
 RL Blochm. Biophys. Res. Commun. 249:56-60(1998).  
 CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS IN ADULT HEART, BRAIN AND  
 CC SKELETAL MUSCLE. LOWER LEVELS IN ADULT PLACENTA, LUNG, KIDNEY AND  
 CC PANCREAS. ALSO HIGH LEVELS IN FETAL BRAIN AND LOWER LEVELS IN  
 CC FETAL LUNG, LIVER AND KIDNEY. ALSO ABUNDANT IN TESTIS AND THYROID.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSION IS INDEPENDENT OF THE CELL CYCLE  
 CC IN LONG FIBROBLASTS.  
 CC -1- SIMILARITY: BELONGS TO THE CYCLIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D50310; BAA08849.1; -;  
 DR EMBL: AF135162; AAF43786.1; -;  
 DR InterPro: IPR000553; -;  
 DR Pfam: PF00134; cyclin; 1;  
 DR PROSITE: PS00292; CYCLINS; FALSE\_NEG.  
 KW Cyclin.  
 FT CONFLICT 9 9 N -> D (IN REF. 3).  
 FT CONFLICT 58 58 Q -> R (IN REF. 3).  
 FT CONFLICT 75 75 R -> G (IN REF. 3).  
 SQ SEQUENCE 377 AA; 42557 MW; 2DE84FEFA74698F6C CRC64;

Query Match 2.0%; Score 7; DB 1; Length 377;  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 SSUPLNS 172  
 Db 252 SSUPLNS 258

RESULT 22  
 PSQA\_ARATH STANDARD; PRT; 386 AA.  
 AC P55034;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, last annotation update)  
 DE 26S PROTEASOME REGULATOR SUBUNIT S5A (MULTIUBQUITIN CHAIN BINDING  
 DE PROTEIN).  
 GN MBP1 OR AT4G38630 OR P20M13.190.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 CC Arabidopsis thaliana (Embryophyta; Tracheophyta; Spermatophyta;  
 CC Magnoliopsida; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 CC Brassicales; Brassicaceae; Arabidopsids.  
 CC NCBI\_TaxId=3702;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA; TISSUE=Seedling;  
 RX MEDLINE=96149398; PubMed=8570648;  
 RA van Nocker S., Deveraux Q., Rechsteiner M., Vierstra R.D.;  
 RA "Arabidopsis MBP1 gene encodes a conserved ubiquitin recognition  
 RA component of the 26S proteasome.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 93:856-860(1996).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=20083488; PubMed=10617198;  
 RA Meyer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,  
 RA Pohl T., Duesterhoeft A., Stiekema K.-D., Terryn N.,  
 RA Harris B., Ansoer W., Brandt P., Grivell L., Rieger M.,  
 RA Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller M.,  
 RA Kreis M., Delsen M., Puigdomenech P., Watson M., Schmidheini T.,  
 RA Reichert B., Foretelle D., Perez-Alonso M., Boutry M., Bancroft I.,  
 RA Vos P., Hohnel J., Zimmermann W., Wedler H., Ridley P.,

RA Langham S.-A., McCullagh B., Bilham L., Robben J., Vandenbussche F.,  
 RA Van der Schueren J., Grymoprez B., Chuang Y.-J., Vandenbussche F.,  
 RA Braeken M., Wellens I., Voet M., Bastiaens I., Aert R., Defoot E.,  
 RA Wetzemager T., Bothe G., Ramsperger U., Hilbert H., Braun M.,  
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirks W.,  
 RA Moijman P., Klein lankhorst R., Rose M., Hauf J., Koetter P.,  
 RA Bernier S., Hempel S., Feldpausch M., Lambert S., Van den Daele H.,  
 RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,  
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,  
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,  
 RA Pettett A., Ralindream M.-A., Lyne M., Benes V., Reckmann S.,  
 RA Borova D., Bloeker H., Scharfe M., Grimm M., Lochert T.-H.,  
 RA Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,  
 RA Neumann S., Arfjou A., Vitale D., Liguori R., Pivrandi E.,  
 RA Massenot O., Quigley F., Chabaud G., Muendlein A., Felber R.,  
 RA Schnabl S., Hiller R., Schmidt W., Decharny A., Abdour S.,  
 RA Chetdor F., Cooke R., Berger C., Montfort A., Casacuberta E.,  
 RA Gibbons T., Weber N., Vandenbol M., Barques M., Terol J., Torres A.,  
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,  
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Blatke C.,  
 RA Fritsman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,  
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,  
 RA Parnell L., Dedhia N., Gnoj L., Schütz K., Huang E., Spiegel L.,  
 RA Senkon M., Murray J., Shee P., Cordes M., Abu-Threideh J.,  
 RA Stoneking T., Kallio J., Graves T., Harmon G., Edwards J.,  
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,  
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,  
 RA Nelson J., Spieth J., Ryan E., Andrews S., Gelsel C., Layman D.,  
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,  
 RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,  
 RA Ma P., Zhong J., Preston R., Wil D., Shekher M., Matero A., Shah R.,  
 RA Saaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,  
 RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,  
 RA Chen E., Marra M., Martienssen R., McCombie W.R.;  
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis  
 RT thaliana.";  
 RL Nature 402:769-777(1999).  
 CC -1- FUNCTION: BINDS AND PRESUMABLY SELECTS UBIQUITIN-CONJUGATES FOR  
 CC DESTRUCTION. PREFERS MULTISUBUNIT CHAINS RATHER THAN SINGLE  
 CC UBIQUITINS.  
 CC -1- SUBUNIT: THE 26S PROTEASOME IS COMPOSED OF A CORE PROTEASE, KNOWN  
 CC AS THE 20S PROTEASOME, CAPPED AT ONE OR BOTH ENDS BY THE 19S  
 CC REGULATORY COMPLEX (RC). THE RC IS COMPOSED OF AT LEAST 18  
 CC DIFFERENT SUBUNITS IN TWO SUBCOMPLEXES, THE BASE AND THE LID,  
 CC WHICH FORM THE PORTIONS PROXIMAL AND DISTAL TO THE 20S PROTEOLYTIC  
 CC CORE, RESPECTIVELY (BY SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: PRESENT IN ALL TISSUES TESTED.  
 CC -1- SIMILARITY: BELONGS TO THE PROTEASOME SUBUNIT S5A FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U33269; AAA85583.1; -;  
 DR EMBL: AL035540; CAB37519.1; -;  
 DR EMBL: AL161593; CAB80527.1; -;  
 KW Proteasome; Repeat.  
 FT DOMAIN 223 318 2 X 12 AA APPROXIMATE REPEATS.  
 FT REPEAT 223 234 1.  
 FT REPEAT 309 318 2.  
 SQ SEQUENCE 386 AA; 40757 MW; 2F5C9D9DFACBA4550 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 386;  
 Best Local Similarity 100.0%; Pred. No. 45;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 SVLSLSP 169  
Db 342 SVLSLSP 348

RESULT 23

Y6B9\_PSEAE STANDARD; PRT; 394 AA.  
ID Y6B9\_PSEAE  
AC P42514;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE HYPOHERICAL PROTEIN PA4219.  
GN PA4219.

OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
[1]

SEQUENCE FROM N.A.  
RC STRAIN-PAOI.  
RX MEDLINE=94117363; PubMed=8288523;  
RA Ankenbauer R.G., Quan H.N.;  
RT "Pptc, the Fe(III)-pyochelin receptor of Pseudomonas aeruginosa: a phenolate siderophore receptor homologous to hydroxamate siderophore receptors.";  
RT J. Bacteriol. 176:307-319(1994).  
[2]

SEQUENCE FROM N.A.  
RC STRAIN-PAOI.  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P., Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M., Garber R.L., Goltzy L., Tolentino E., Westbrook-Madman S., Yuan Y., Brody L.L., Coulter S.N., Folger K.R., Kas A., Laidig K., Lim R.M., Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T., Ralizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RA "Complete genome sequence of Pseudomonas aeruginosa PAOI, an opportunistic pathogen.";  
RT Nature 400:595-964(2000).  
[1] SUPRACELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; 003161; AAC43215.1; -;  
DR EMBL; AE004838; AAG07607.1; -;  
KM Hypothetical protein; Transmembrane.  
FT TRANSMEM 22 42  
FT TRANSMEM 60 80  
FT TRANSMEM 81 101  
FT TRANSMEM 231 251  
FT TRANSMEM 271 291  
FT TRANSMEM 303 323  
FT TRANSMEM 328 348  
FT TRANSMEM 355 375  
SO SEQUENCE 394 AA; 42248 MW; 89AD093871CC71BE CRC64;

Query Match 2.0%; Score 7; DB 1; Length 394;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 IQLAPSE 131  
Db 287 IQLAPSE 293

RESULT 24  
CN37\_RAT STANDARD; PRT; 404 AA.  
ID CN37\_RAT  
AC P13233;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE 2',3'-CYCLIC NUCLEOTIDE 3'-PHOSPHODIESTERASE (EC 3.1.4.37) (CNP)  
GN CNP.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
[1]

SEQUENCE FROM N.A.  
RX MEDLINE=87310616; PubMed=3040924;  
RA Bernier L., Alvarez F., Norgard E.M., Raible D.W., Mentaberry A., Schembri J.G., Sabatini D.D., Colman D.R.;  
RT "Molecular cloning of a 2',3'-cyclic nucleotide 3'-phosphodiesterase: mRNAs with different 5' ends encode the same set of proteins in nervous and lymphoid tissues.";  
RT J. Neurosci. 7:2703-2710(1987).  
CC -1- CATALYTIC ACTIVITY: NUCLEOSIDE 2',3'-CYCLIC PHOSPHATE + H(2)O = -1- NUCLEOSIDE 2'-PHOSPHATE.  
CC -1- SUPRACELLULAR LOCATION: FIRMLY BOUND TO MEMBRANE STRUCTURES OF BRAIN WHITE MATTER.  
[2]

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M18630; AAA40939.1; -;  
DR PIR; A45670; A45670.  
KM Hydrolyase; Membrane; Brain.  
SO SEQUENCE 404 AA; 45595 MW; 9C4CA888E693BFF CRC64;

Query Match 2.0%; Score 7; DB 1; Length 404;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 267 KSEFDSL 273  
Db 205 KSEFDSL 211

RESULT 25  
FABL\_MYCTU STANDARD; PRT; 416 AA.  
ID FABL\_MYCTU  
AC Q10524;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE PUTATIVE 3-OXOACYL-LACYL-CARRIER-PROTEIN SYNTHASE 1 (EC 2.3.1.41)  
DE (BETA-KETOACYL-ACP SYNTHASE 1) (KAS 1).  
GN RV2245 OR MYCVA27.26.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
[1]

SEQUENCE FROM N.A.  
RC STRAIN-H37Rv;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garman T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,

RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Sgares R., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 CC -1- FUNCTION: CATALYZES THE CONDENSATION REACTION OF FATTY ACID  
 CC SYNTHESIS BY THE ADDITION TO AN ACYL ACCEPTOR OF TWO CARBONS  
 CC FROM MALONYL-ACP (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ACYL-[ACYL-CARRIER PROTEIN] + MALONYL-[ACYL-  
 CC CARRIER PROTEIN] = 3-OXOACYL-[ACYL-CARRIER PROTEIN] + CO(2) +  
 CC [ACYL-CARRIER PROTEIN].  
 CC -1- PATHWAY: FATTY ACID BIOSYNTHESIS.  
 CC -1- SUBCELLULAR LOCATION: CYTOSOL; PLASMA MEMBRANE (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO THE BETA-KETOACYL-ACP SYNTHETASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: 270692; CA94641.1; -.  
 CC Tuberculolysin: RV2245; -.  
 CC InterPro: IPR000794; -.  
 CC Pfam: PF00109; ketoacyl-synt. 1.  
 CC PROSITE: PS00606; B. KETOACYL SYNTHETASE; FALSE NEG.  
 CC Hypothetical protein; Fatty acid biosynthesis; Transferase;  
 CC Acyltransferase.  
 CC ACYL SITE 171 BY SIMILARITY.  
 CC FT ACYL SITE 171 MW: D2187BE2F0B56C7F CRC64;  
 CC S0 SEQUENCE 416 AA; 43316 MW; D2187BE2F0B56C7F CRC64;  
 CC  
 CC Query Match 2.0%; Score 7; DB 1; Length 416;  
 CC Best Local Similarity 100.0%; Pred. No. 48;  
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 213 SLEAGL 219  
 CC |||||  
 CC Db 294 SLEAGL 300  
 CC  
 CC RESULT 26  
 CC MTDL\_NEMC STANDARD: PRT: 420 AA.  
 CC ID MTDL\_NEMC  
 CC AC G9RLM4;  
 CC DT 01-OCT-2000 (Rel. 40, Created)  
 CC DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 CC DE 01-OCT-2000 (Rel. 40, Last annotation update)  
 CC DE PROBABLE MODIFICATION METHYLASE NMEIP (EC 2.1.1.73) (CYTOSINE-  
 CC SPECIFIC METHYLTRANSFERASE NMEIP) (M.NMEIP).  
 CC GN NMEIP.  
 CC OS Neisseria meningitidis (serogroup C).  
 CC CC Bacteria; Proteobacteria; Beta subdivision; Neisseriaceae; Neisseria.  
 CC NCBI\_TaxID=135720;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC RC STRAIN-2120 / SEROGROUP C / SEROTYPE NT;  
 CC RX MEDLINE=20138154; PubMed=10671450;  
 CC RA Claus H., Friedrich A., Frosch M., Vogel U.;  
 CC RT "Differential distribution of novel restriction-modification systems  
 CC in clonal lineages of Neisseria meningitidis.";  
 CC RL J. Bacteriol. 182:1296-1303(2000).  
 CC -1- FUNCTION: THIS METHYLASE RECOGNIZES A DOUBLE-STRANDED SEQUENCE,  
 CC CAUSES SPECIFIC METHYLATION ON A C ON BOTH STRANDS, AND PROTECTS  
 CC THE DNA FROM CLEAVAGE BY THE NMEIP ENDONUCLEASE (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + DNA CYTOSINE =  
 CC S-ADENOSYL-L-HOMOCYSTEINE + DNA 5-METHYLCYTOSINE.  
 CC -1- SIMILARITY: STRONG. TO OTHER C5-DNA METHYLASES.  
 CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: A238948; CAB59897.1; -.  
 CC HSP: P20589; IDCT.  
 CC DR REBASE: 4188; M.NMEIP.  
 CC DR InterPro: IPR001525; -.  
 CC DR Pfam: PF00145; DNA methylase; 2.  
 CC DR PRINTS: PF00105; C5METHYLASE.  
 CC DR PROSITE: PS00094; C5\_MTHASE.1; FALSE NEG.  
 CC DR PROSITE: PS00095; C5\_MTHASE.2; FALSE NEG.  
 CC KW Transferase; Methyltransferase; Restriction system.  
 CC FT ACYL SITE 148 148 BY SIMILARITY.  
 CC FT ACYL SITE 148 148 MW: 6E2472BA070354C9 CRC64;  
 CC S0 SEQUENCE 420 AA; 47930 MW; 6E2472BA070354C9 CRC64;  
 CC  
 CC Query Match 2.0%; Score 7; DB 1; Length 420;  
 CC Best Local Similarity 100.0%; Pred. No. 48;  
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 241 VDALKF 247  
 CC |||||  
 CC Db 244 VDALKF 250  
 CC  
 CC RESULT 27  
 CC SLSL\_YARLI STANDARD: PRT: 426 AA.  
 CC ID SLSL\_YARLI  
 CC AC G99158;  
 CC DT 01-NOV-1997 (Rel. 35, Created)  
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 CC DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 CC DE SLSL PROTEIN PRECURSOR.  
 CC GN SLSL.  
 CC OS Yarrowia lipolytica (Candida lipolytica).  
 CC CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC CC Saccharomycetales; Dipodascaceae; Yarrowia.  
 CC NCBI\_TaxID=4952;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC RC STRAIN-ATCC 20460 / W29;  
 CC RX MEDLINE=96216076; PubMed=8662639;  
 CC RA Boistrame A., Beckerich J.-M., Galliard C.;  
 CC RT "Sls1, an endoplasmic reticulum component, is involved in the  
 CC protein translocation process in the yeast Yarrowia lipolytica.";  
 CC RL J. Biol. Chem. 271:11668-11675(1996).  
 CC -1- FUNCTION: INVOLVED IN THE PROTEIN TRANSLOCATION PROCESS. MAY  
 CC INTERACT DIRECTLY WITH TRANSLOCATING POLYPEPTIDES TO FACILITATE  
 CC THEIR TRANSFER AND/OR HELP THEIR FOLDING IN THE ER. IT IS NOT  
 CC REQUIRED FOR VIABILITY BUT IS ESSENTIAL FOR OPTIMAL GROWTH AT  
 CC ELEVATED TEMPERATURES.  
 CC -1- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: 250154; CA90516.1; -.  
 CC InterPro: IPR000886; -.  
 CC DR PROSITE: PS00014; ER\_TARGET.1.  
 CC KW Endoplasmic reticulum; signal.  
 CC FT SIGNAL 1 17  
 CC FT CHAIN 18 426  
 CC FT SLSL PROTEIN.  
 CC FT SITE 423 426 PREVENT SECRETION FROM ER (POTENTIAL).  
 CC

SEQUENCE 426 AA; 47201 MW; 0ACD7EF17540B8E2 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 426;

Best Local Similarity 100.0%; Pred. No. 49;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LALVAG 13

DB 8 LALVAG 14

RESULT 28

IRF4\_MOUSE STANDARD; PRT; 450 AA.

AC 064287; Q60802;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE INTERFERON REGULATORY FACTOR 4 (IRF-4) (LYMPHOCYTE SPECIFIC INTERFERON

REGULATORY FACTOR) (ISIR) (NF-BM5) (PU.1 INTERACTION PARTNER)

DE (TRANSCRIPTIONAL ACTIVATOR PTP).

IRF4 OR SPIP.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

[1]

RP SEQUENCE FROM N.A.

RC STRAIN=BAIB/C;

RA MEDLINE=95317607; PubMed=7797077;

RA Eisenbeis C.F., Singh H., Stoch U.;

RT "Pip, a novel IRF family member, is a lymphoid-specific,

RT PU.1-dependent transcriptional activator.";

RL Genes Dev. 9:1377-1387(1995).

[2]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6; AND 129/SVJ; TISSUE=Spleen;

RA MEDLINE=95334364; PubMed=7541907;

RA Matsuyama T., Grossman A., Mitruecker H.-W., Siderovski D.P.,

RA Kleifer F., Kawakami T., Richardson C.D., Taniguchi T., Yoshinaga S.K.,

RA Mak T.W.;

RT "Molecular cloning of ISIRF, a lymphoid-specific member of the

RT interferon regulatory factor family that binds the interferon-

RT Nucleic Acids Res. 23:2127-2136(1995).

CC -I- FUNCTION: TRANSCRIPTIONAL ACTIVATOR. BINDS TO THE INTERFERON-

CC STIMULATED RESPONSE ELEMENT (ISRE) OF THE MHC CLASS I PROMOTER.

CC BINDS THE IMMUNOGLOBULIN LAMBDA LIGHT CHAIN ENHANCER, TOGETHER

CC WITH PU.1. PROBABLY PLAYS A ROLE IN ISRE-TARGETED SIGNAL

CC TRANSDUCTION MECHANISMS SPECIFIC TO LYMPHOID CELLS.

CC -I- SUBCELLULAR LOCATION: NUCLEAR.

CC -I- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1 (SHOWN HERE) AND 2; ARE

CC PRODUCED BY ALTERNATIVE SPLICING. THEY ONLY DIFFER IN THE PRESENCE

CC OR ABSENCE OF A GLUTAMIC ACID RESIDUE.

CC -I- TISSUE SPECIFICITY: LYMPHOID CELLS.

CC -I- INDUCTION: NOT INDUCED BY INTERFERONS.

CC -I- SIMILARITY: BELONGS TO THE IRF FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

CC EMBL: U34307; AAA75283.1; -

DR EMBL: U11692; AAA75309.1; -

DR EMBL: U20949; AAA75316.1; -

DR EMBL: U20949; AAA75317.1; -

DR HSSP: P23906; IIRG.

DR MGD: MGI:1096873; Irf4.

DR MGD: MGI:104895; SPIP.

DR InterPro: IPR001346; -

DR Pfam: PF00605; IRF; 1.

DR PRINTS: PR00267; INTERNEGCT.

DR PROSITE: PS00601; IRF; 1.

KW Transcription regulation; DNA-binding; Nuclear protein; Activator;

KW Alternative splicing.

FT DNA\_BIND 23 125 TRYPTOPHAN PENTAD REPEAT.

FT VARSPIC 165 165 MISSING (IN ISOFORM 2).

FT VARSPIC 165 165

QY SEQUENCE 450 AA; 51577 MW; 5FD94CA6C453869C CRC64;

Query Match

Best Local Similarity 100.0%; Pred. No. 51;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 FUSELOV 189

DB 364 FUSELOV 370

RESULT 29

YD36\_YEAST STANDARD; PRT; 462 AA.

AC 012069;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE HYPOTHETICAL 53.4 KDA PROTEIN IN PRP9-NAT1 INTERGENIC REGION.

GN YD1036C OR D2743.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OX NCBI\_TaxID=4932;

[1]

RP SEQUENCE FROM N.A.

RC STRAIN=S288C;

RA Paulin L., Saren A.M., Laamanen P.;

RA Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.

CC -I- SIMILARITY: BELONGS TO THE RLU FAMILY OF PSEUDOURIDINE SYNTHASES.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

CC EMBL: Z71781; CAA96453.1; -

DR EMBL: Z74084; CAA98595.1; -

DR SGD: S0002194; YD1036C.

DR InterPro: IPR000613; -

DR InterPro: IPR002990; -

DR Pfam: PF00849; Pseudou\_synth.2; 1.

DR PROSITE: PS01129; PSI\_RLU; 1.

KW Hypothetical protein.

QY SEQUENCE 462 AA; 53399 MW; 300E676D8E3819F0 CRC64;

DB 382 LDEIGKR 388

QY 219 LDEIGKR 225

DB 382 LDEIGKR 388

Query Match

Best Local Similarity 100.0%; Pred. No. 52;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 219 LDEIGKR 225

DB 382 LDEIGKR 388

RESULT 30

PROD\_HUMAN STANDARD; PRT; 516 AA.

ID PROD\_HUMAN

AC 043272;

```

DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, last sequence update)
DT 01-OCT-2000 (Rel. 40, last annotation update)
DE PROLINE OXIDASE, MITOCHONDRIAL PRECURSOR (EC 1.5.3.-) (PROLINE
DE DEHYDROGENASE).
EN PROD#..
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain:
RX MEDLINE=98046348; PubMed=9385373;
RA Campbell H.D., Webb G.C., Young I.G.;
RA "A human homologue of the Drosophila melanogaster slungish-A (proline
RT oxidase) gene maps to 22q11.2, and is a candidate gene for type-I
RT hyperproliferemia."
RL Hum. Genet. 101:69-74(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum, and Kidney:
RX MEDLINE=99206616; PubMed=10192398;
RA Gogos J.A., Santha M., Takacs Z., Beck K.D., Luine V., Lucas L.R.,
RA Nadler J.V., Karayiorgou M.;
RT "The gene encoding proline dehydrogenase modulates sensorimotor gating
RT in mice";
RL Nat. Genet. 21:434-439(1999).
CC -1- FUNCTION: CONVERTS PROLINE TO DELTA-1-PYRROLINE-5-CARBOXYLATE.
CC -1- PATHWAY: FIRST STEP IN THE CONVERSION FROM PROLINE TO GLUTAMATE.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN LUNG, SKELETAL MUSCLE AND BRAIN,
CC TO A LESSER EXTENT IN HEART AND KIDNEY, AND WEAKLY IN LIVER,
CC PLACENTA AND PANCREAS.
CC -1- DISEASE: DEFECTS IN PRODH ARE A CAUSE OF TYPE-1 HYPERPROLINEMIA.
CC A DISORDER WHERE SERUM PROLINE LEVELS ARE ELEVATED. MAY BE
CC INVOLVED IN THE 22Q11 ASSOCIATED PSYCHIATRIC AND BEHAVIORAL
CC PHENOTYPES.
CC -1- SIMILARITY: BELONGS TO THE PROLINE OXIDASE FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its use
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U82381; AAB88789.1; -.
DR EMBL; AF120278; AAD24775.1; -.
DR MIM; 239500; -.
DR InterPro: IPR002872; -.
DR Pfam: PF01619; Pro.dh; 1.
KW oxidoreductase; Proline metabolism; Mitochondrion; Transl. peptid.
FT TRANSIT 1 ?
FT CHAIN 80 516
FT CONFLICT 60 80
FT CONFLICT 80 80
FT TRANSIT 1 ?
FT CHAIN 80 516
FT CONFLICT 60 80
FT CONFLICT 80 80
SQ SEQUENCE 516 AA; 59216 MW; 2FA5B1E4481C450A CRC64;
Query Match 2.0%; Score 7; DB 1; Length 516;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 66 GNLEHRP 72
Db 509 GNLEHRP 515
RESULT 31
CSTA_AQAAE STANBARD: PRT: 555 AA.
ID CSTA_AQAAE STANBARD: PRT: 555 AA.
AC 067304;

```

DR	EMBL: AF000731; AAC07258.1; -.	CC	-----	CC	
KM	Transmembrane. 1	FT	21	POTENTIAL.	
FT	TRANSMEM	FT	55	POTENTIAL.	
FT	TRANSMEM	FT	75	POTENTIAL.	
FT	TRANSMEM	FT	76	POTENTIAL.	
FT	TRANSMEM	FT	131	POTENTIAL.	
FT	TRANSMEM	FT	151	POTENTIAL.	
FT	TRANSMEM	FT	160	POTENTIAL.	
FT	TRANSMEM	FT	180	POTENTIAL.	
FT	TRANSMEM	FT	186	POTENTIAL.	
FT	TRANSMEM	FT	225	POTENTIAL.	
FT	TRANSMEM	FT	245	POTENTIAL.	
FT	TRANSMEM	FT	251	POTENTIAL.	
FT	TRANSMEM	FT	271	POTENTIAL.	
FT	TRANSMEM	FT	290	POTENTIAL.	
FT	TRANSMEM	FT	310	POTENTIAL.	
FT	TRANSMEM	FT	327	POTENTIAL.	
FT	TRANSMEM	FT	347	POTENTIAL.	
FT	TRANSMEM	FT	388	POTENTIAL.	
FT	TRANSMEM	FT	433	POTENTIAL.	
FT	TRANSMEM	FT	453	POTENTIAL.	
FT	TRANSMEM	FT	462	POTENTIAL.	
FT	TRANSMEM	FT	482	POTENTIAL.	
FT	TRANSMEM	FT	494	POTENTIAL.	
FT	TRANSMEM	FT	514	POTENTIAL.	
FT	TRANSMEM	FT	523	POTENTIAL.	
FT	TRANSMEM	FT	543	POTENTIAL.	
SO	SOURCE	555 AA; 60113 MW; B3BC690C460E/F04 CRC64;			
Query Match					
Best Local Similarity 2.0%; Score 7; DB 1; Length 555;					
Matches 7; Conservative 0; Pred. No. 61;					
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
OY	317 LALAVII 323				
DB	339 LALAVII 345				
RESULT 32					
TPPL_MOUSE	STANDARD;	PRT;	562 AA.		
AC	089023; O90057;				
DT	15-DEC-1998 (Rel. 37, Created)				
DT	01-OCT-2000 (Rel. 40, Last sequence update)				
DT	01-OCT-2000 (Rel. 40, Last annotation update)				
DE	TRIPPEPTIDYL-PEPTIDASE 1 PRECURSOR (EC 3.4.14.9) (TPP-I) (TRIPPEPTIDYL				
DE	AMINOPEPTIDASE) (LYSOSOMAL PEPTASTIN INSENSITIVE PROTEASE)				
DE	(LPTC).				
GN	CLN2 OR TPPL.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Rodentia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				

[1] SEQUENCE FROM N.A.  
 RP MEDLINE-99142696; PubMed-9989590;  
 RA Vines D.J., Warburton M.J.;  
 RT "Classical late infantile neuronal ceroid lipofuscinosis fibroblasts  
 are deficient in lysosomal tripeptidyl peptidase I.";  
 RL FEBS Lett. 443:131-135(1999).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-20028316; PubMed-10556422;  
 RA Katz M.L., Liu P.-C., Grob-Numm S.F., Shbuaya H., Johnson G.S.;  
 RT "Characterization and chromosomal mapping of a mouse ortholog of the  
 late-infantile ceroid-lipofuscinosis gene CLN2.";  
 RL Mamm. Genome 10:1050-1053(1999).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RA Steat D.E., Lobel P.;  
 RT "Murine homologue of the lysosomal pepstatin insensitive protease  
 which is deficient in human classical late infantile neuronal ceroid  
 lipofuscinosis.";  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: RELEASE OF AN N-TERMINAL TRIPEPTIDE FROM A  
 POLYPEPTIDE.  
 CC -1- SUBCELLULAR LOCATION: LYSOSOMAL.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S53.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AJ011912; CAA03863.1; ALT\_INIT.  
 DR EMBL: AF124599; AAD32573.1; -  
 DR EMBL: AF111172; AAD03083.1; -  
 DR MEROPS: S53.003; -  
 KM MGD: MGI:1336194; CLN2.  
 KM Hydrolyase, Protease; Serine protease; Zymogen; Signal; Lysosome;  
 KM Glycoprotein.  
 FT SIGNAL 1 16 POTENTIAL.  
 FT PROPEP 17 194 REMOVED IN MATURE FORM (BY SIMILARITY).  
 FT CHAIN 195 562 TRIPEPTIDYL-PEPTIDASE I.  
 FT CARBOHD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHD 221 221 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHD 312 312 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHD 442 442 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT ACT\_SITE 359 359 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT ACT\_SITE 516 516 BY SIMILARITY.  
 FT CONFLICT 1 1 M -> V (IN REF. 1).  
 FT CONFLICT 562 562 P -> LDPVP (IN REF. 1).  
 SO SEQUENCE 562 AA; 61362 MW; 0AF8163E8A1A66396 CRC64;  
 Query Match 2.0%; Score 7; DB 1; Length 562;  
 Best Local Similarity 100.0%; Pred. No. 62;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 43 PDVAALS 49  
 | | | | |  
 Db 449 PDVAALS 455  
 RESULT 33  
 TPPL\_HUMAN STANDARD; PRT; 563 AA.  
 AC 014773;  
 DT 15-DEC-1998 (Rel. 37; Created)  
 DT 15-DEC-1998 (Rel. 37; Last sequence update)  
 DT 01-OCT-2000 (Rel. 40; Last annotation update)  
 DE TRIPEPTIDYL-PEPTIDASE I PRECURSOR (EC 3.4.14.9) (TPP-1) (TRIPEPTIDYL

DE AMINOPEPTIDASE) (LYSOSOMAL PEPSTATIN INSENSITIVE PROTEASE)  
 GN (LIPC).  
 OS CLN2 OR TPPL.  
 OC Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND VARIANTS LINC1 R-365; Y-365.  
 RC TISSUE=Placenta;  
 RX MEDLINE-97442529; PubMed-9295367;  
 RA Steat D.E., Donnelly R.J., Lackland H., Liu C.-G., Sohar I.,  
 RA Pullarkat R.K., Lobel P.;  
 RT "Association of mutations in a lysosomal protein with classical  
 late-infantile neuronal ceroid lipofuscinosis.";  
 RL Science 277:1802-1805(1997).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE-98317534; PubMed-9653647;  
 RA Liu C.-G., Steat D.E., Donnelly R.J., Lobel P.;  
 RT "Structural organization and sequence of CLN2, the defective gene is  
 classical late infantile neuronal ceroid lipofuscinosis.";  
 RL Genomics 50:206-212(1998).  
 [3]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE-99406495; PubMed-10477428;  
 RA Mole S.E., Mitchelson H.M., Munroe P.B.;  
 RT "Molecular basis of the neuronal ceroid lipofuscinoses: mutations in  
 CLN1, CLN2, CLN3, and CLN5.";  
 RL Hum. Mutat. 14:199-215(1999).  
 [4]  
 RP VARIANTS LINC1.  
 RX MEDLINE-99264233; PubMed-10330339;  
 RA Steat D.E., Gin R.M., Sohar I., Wisniewski K., Sklower-Brooks S.,  
 RA Pullarkat R.K., Palmer D.N., Lerner T.J., Boustany R.M., Uldall P.,  
 RA Shkarots A.N., Donnelly R.J., Lobel P.;  
 RT "Mutational analysis of the defective protease in classic  
 late-infantile neuronal ceroid lipofuscinosis, a neurodegenerative  
 lysosomal storage disorder.";  
 RL Am. J. Hum. Genet. 64:1511-1523(1999).  
 CC -1- CATALYTIC ACTIVITY: RELEASE OF AN N-TERMINAL TRIPEPTIDE FROM A  
 POLYPEPTIDE.  
 CC -1- SUBCELLULAR LOCATION: LYSOSOMAL.  
 CC -1- TISSUE SPECIFICITY: DETECTED IN ALL TISSUES EXAMINED WITH HIGHEST  
 LEVELS IN HEART AND PLACENTA AND RELATIVELY SIMILAR LEVELS IN  
 OTHER TISSUES.  
 CC -1- DISEASE: DEFECTS IN CLN2 ARE THE CAUSE OF CLASSICAL LATE-INFANTILE  
 NEURONAL CEROID LIPOFUSCINOSIS (LINC1) (ALSO KNOWN AS CEROID  
 LIPOFUSCINOSIS NEURONAL 2; CLN2); A FATAL CHILDHOOD  
 NEURODEGENERATIVE DISEASE CHARACTERIZED BY PROGRESSIVE VISUAL AND  
 MENTAL DECLINE, MOTOR DISTURBANCE, EPILEPSY AND BEHAVIORAL  
 CHANGES. THE THREE MAIN SUBTYPES OF CHILDHOOD NCLs DEFINED BY THE  
 AGE OF ONSET, CLINICAL FEATURES, AND ULTRASTRUCTURAL MORPHOLOGY  
 ARE INFANTILE NCL (INCL), CLASSICAL LATE-INFANTILE NCL (LINC1), OR  
 JUVENILE NCL (JNCL), ALTHOUGH A NUMBER OF OTHER DISTINCT VARIANTS  
 FORMS HAVE BEEN DESCRIBED.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S53.  
 CC -1- DATABASE: NAME=NCL CLN2;  
 CC NOTE=Neural Ceroid Lipofuscinoses mutation db;  
 CC WWW="http://www.ucl.ac.uk/ncl/CLN2.html".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF017456; AAB80725.1; -  
 DR EMBL: AF039704; AAC98480.1; -  
 DR MEROPS: S53.003; -



DR MIM; 204500; -  
 KM Hydrolase; Protease; Serine protease; Zymogen; Signal; Lysosome;  
 KM Glycoprotein; Neuronal ceroid lipofuscinosis; Disease mutation;  
 KM Polymorphism.

FT SIGNAL 1 16  
 FT PROPEP 17 195  
 FT CHAIN 196 363  
 FT CAROHND 210 210  
 FT CAROHND 222 222  
 FT CAROHND 286 286  
 FT CAROHND 313 313  
 FT CAROHND 443 443  
 FT ACT\_SITE 360 360  
 FT ACT\_SITE 517 517  
 FT ACT\_SITE 77 77  
 FT VARIANT 100 100  
 FT VARIANT 175 175  
 FT VARIANT 206 206  
 FT VARIANT 287 287  
 FT VARIANT 343 343  
 FT VARIANT 365 365  
 FT VARIANT 365 365  
 FT VARIANT 385 385  
 FT VARIANT 389 389  
 FT VARIANT 422 422  
 FT VARIANT 447 447  
 FT VARIANT 454 454  
 FT VARIANT 475 475  
 FT SEQUENCE 563 AA; 61228 MW; 75CA68792EC75647 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 563;  
 Best Local Similarity 100.0%; Pred. No. 62;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 43 PDVALS 49  
 DB 450 PDVALS 456

RESULT 34  
 MUTL\_CHLTR STANDARD; PRT; 576 AA.  
 ID 084579;  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE DNA MISMATCH REPAIR PROTEIN MUTL.  
 GN MUTL OR CT575.  
 OS Chlamydia trachomatis.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_Taxid=813;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=D/UM-3/CX;  
 RA MEDLINE=99000809; PubMed=9784136;  
 RA Stephens R.S., Kaiman S., Lammell C.J., Fan J., Marathe R., Aravind L.,

RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
 RA Davis R.W.;  
 RT "Genome sequence of an obligate intracellular pathogen of humans:  
 RT Chlamydia trachomatis".  
 RL Science 282:754-759(1998).  
 CC -1- FUNCTION: THIS PROTEIN IS INVOLVED IN THE REPAIR OF MISMATCHES IN  
 CC DNA. IT IS REQUIRED FOR DAM-DEPENDENT METHYL-DIRECTED DNA MISMATCH  
 CC REPAIR. MAY ACT AS A "MOLECULAR MATCHMAKER", A PROTEIN THAT  
 CC PROMOTES THE FORMATION OF A STABLE COMPLEX BETWEEN TWO OR MORE  
 CC DNA-BINDING PROTEINS IN AN ATP-DEPENDENT MANNER WITHOUT ITSELF  
 CC BEING PART OF A FINAL EFFECTOR COMPLEX (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTL/HEXV FAMILY.  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AE001328; AAC68177.1; -  
 CC DR Pfam: PF01119; DNA\_mis\_repair; 1.  
 CC DR PROSITE: PS00058; DNA\_MISMATCH\_REPAIR\_1; 1.  
 CC KM DNA repair; 576 AA; 64138 MW; 8A6FCCL19783F84E CRC64;  
 CC SEQUENCE

Query Match 2.0%; Score 7; DB 1; Length 576;  
 Best Local Similarity 100.0%; Pred. No. 63;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 161 ENSVLS 167  
 DB 179 ENSVLS 185

RESULT 35  
 SYFB\_DROME STANDARD; PRT; 589 AA.  
 ID 03VC45;  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE PROBABLE PHENYLALANINE-TRNA SYNTHETASE BETA CHAIN (EC 6.1.1.20)  
 DE (PHENYLALANINE--TRNA LIGASE BETA CHAIN) (PHERS).  
 GN CG5706.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_Taxid=7227;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RC MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Gelniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Planck C., Baldwin D.,  
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Bens P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merklov G., Mishina N.Y., Mobarry C., Morris J., Moshell A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy U., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Sliden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Wolley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 CC -1- CATALYTIC ACTIVITY: ATP + L-PHENYLTALANINE + TRNA(PHE) = AMP +  
 CC PYROPHOSPHATE + L-PHENYLTALANIL-TRNA(PHE).  
 CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (by  
 CC SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE PHENYLTALANIL-TRNA SYNTHETASE BETA  
 CC CHAIN FAMILY. SUBFAMILY 2.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: AE003747; AAF56268.1; -  
 DR EMBL: FB90039175; CG5706.  
 KM Aminoacyl-tRNA synthetase; Protein biosynthesis; Liasase; ATP-binding.  
 SQ SEQUENCE 589 AA: 65777 MW: D7596D94E97AADA1 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 589;  
 Best Local Similarity 100.0%; Pred. No. 65;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 91 GSVISYP 97  
 | | | | | | | | | |  
 Db 294 GSVISYP 300

RESULT 36  
 ID GHR\_RABIT STANDARD: PRT: 638 AA.  
 AC P19941;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, last sequence update)  
 DT 15-JUL-1998 (Rel. 36, last annotation update)  
 DE GROWTH HORMONE RECEPTOR PRECURSOR (GH RECEPTOR) (SERUM BINDING  
 DE PROTEIN).  
 GN GHR.  
 OS *Oryctolagus cuniculus* (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=8605896; PubMed=2825030;  
 RA Leung D.W., Spencer S.A., Cachianes G., Hammonds R.G., Collins C.,  
 RA Henz W.J., Barnard R., Waters M.J., Wood W.I.;  
 RT "Growth hormone receptor and serum binding protein: purification,  
 RT cloning and expression.";

RL Nature 330:537-543(1987).  
 CC -1- FUNCTION: THIS IS A RECEPTOR FOR PITUITARY GLAND GROWTH HORMONE.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.  
 CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: AF015252; AAB67613.1; -  
 DR PIR: S08544; S08544.  
 DR HSSP: P10912; 1A22.  
 DR InterPro: IPR000950; -  
 DR InterPro: IPR001777; -  
 DR InterPro: IPR002465; -  
 DR Pfam: PF00041; Fn3; 1.  
 DR PROSITE: PS01352; HEMATOPOY\_REC\_L\_F1; 1.  
 KW Receptor; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 18  
 FT CHAIN 19 638  
 FT DOMAIN 19 264  
 FT TRANSMEM 265 288  
 FT DOMAIN 289 638  
 FT DISULFID 145 252  
 FT DISULFID 56 66  
 FT DISULFID 101 112  
 FT DISULFID 126 140  
 FT CARBOHYD 46 46  
 FT CARBOHYD 115 115  
 FT CARBOHYD 156 156  
 FT CARBOHYD 161 161  
 FT CARBOHYD 200 200  
 SQ SEQUENCE 638 AA: 71076 MW: E05CCE1D7294624C CRC64;

Query Match 2.0%; Score 7; DB 1; Length 638;  
 Best Local Similarity 100.0%; Pred. No. 69;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 22 LKSPGSV 28  
 | | | | | | | | | |  
 Db 76 LKSPGSV 82

RESULT 37  
 ID S3A1\_HUMAN STANDARD: PRT: 793 AA.  
 AC Q15459;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, last sequence update)  
 DT 01-OCT-2000 (Rel. 40, last annotation update)  
 DE SPLICING FACTOR 3 SUBUNIT 1 (SPLICOSOME ASSOCIATED PROTEIN 114) (SAP  
 DE 114) (SF3A120).  
 GN SF3A1 OR SAP114.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96079958; PubMed=7489498;  
 RA Kramer A., Mulhauser F., Wersig C., Groning R., Bilbe G.;  
 RT "Mammalian splicing factor SF3A120 represents a new member of the  
 RT SUP family of proteins and is homologous to the essential splicing  
 RT factor PRP23 of *Saccharomyces cerevisiae*.";  
 RN RNA 1:260-272(1995).  
 [2]

RP SEQUENCE FROM N.A.  
RA Bentley D., Blandford M.;  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: SUBUNIT OF THE SF3A COMPLEX REQUIRED FOR 'A' COMPLEX  
CC ASSEMBLY FORMED BY THE STABLE BINDING OF U2 SNRP TO THE  
CC BRANCHPOINT SEQUENCE (BPS) IN PRE-MRNA. SEQUENCE INDEPENDENT  
CC BINDING OF SF3A/SF3B COMPLEX UPSTREAM OF THE BRANCH SITE IS  
CC ESSENTIAL, IT MAY ANCHOR U2 SNRP TO THE PRE-MRNA.  
CC -1- SUBUNIT: SF3A IS COMPOSED OF SAYS 61, 62 AND 114 (SF3A60, SF3A66  
CC AND SF3A120 RESPECTIVELY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
CC -1- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED.  
CC -1- SIMILARITY: CONTAINS 1 UBQUITIN-LIKE DOMAIN.  
CC -1- SIMILARITY: TO YEAST PRE-MRNA SPLICING FACTOR PRP21.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
CC -----  
DR EMBL: X85237; CA839494.1; -;  
DR EMBL: AC004997; AAC23435.1; -;  
DR InterPro: IPR000061; -;  
DR InterPro: IPR000626; -;  
DR Pfam: PF01805; Surp; 2.  
DR Pfam: PF00240; ubiquitin; 1.  
DR PROSITE: PS50053; UBQUITIN\_2; 1.  
DR mRNA processing; mRNA splicing; Nuclear protein; Repeat.  
KM DOMAIN 10 16 POLY-PRO.  
FT DOMAIN 118 122 POLY-GLN.  
FT DOMAIN 260 267 POLY-GLU.  
FT DOMAIN 369 372 POLY-PRO.  
FT DOMAIN 557 560 POLY-PRO.  
FT DOMAIN 672 675 POLY-PRO.  
FT DOMAIN 707 793 UBQUITIN-LIKE.  
FT DOMAIN 793 AA; 88886 MW; 7259F1EC4577305C CRC64;  
SQ SEQUENCE 793 AA; 88886 MW; 7259F1EC4577305C CRC64;  
  
Query Match 2.0%; Score 7; DB 1; Length 793;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RP SEQUENCE FROM N.A.  
RX MEDLINE:97245296; PubMed:9090055;  
RA de Antoni A., D'Angelo M., Dal Pero F., Sartorello F., Pandolfo D.,  
RA Pallavicini A., Lanfranchi G., Valle G.;  
RT "The DNA sequence of cosmid 14-13b from chromosome XIV of  
RT Saccharomyces cerevisiae reveals an unusually high number of  
RT overlapping open reading frames.";  
RL Yeast 13:261-266(1997).  
RN [3]  
RP CHARACTERIZATION.  
RX MEDLINE:96324398; PubMed:8670895;  
RA Geisler S., Pereira G., Spang A., Knop M., Soues S., Kilmarlin J.V.,  
RA Scheibel E.;  
RT "The spindle pole body component Spc98p interacts with the  
RT gamma-tubulin-like Tub4p of Saccharomyces cerevisiae at the sites of  
RT microtubule attachment";  
RL EMBL J.15:389-391(1996)  
CC -1- FUNCTION: INTERACTS WITH TUB4 TO FORM A COMPLEX INVOLVED IN  
CC MICROTUBULE ORGANIZATION BY THE SPINDLE POLE BODY (SPB).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
CC -----  
DR EMBL: Z46843; CA868899.1; -;  
DR EMBL: Z69382; CA93378.1; -;  
DR EMBL: Z71402; CA96007.1; -;  
DR SCD: S0005070; SPC98.  
KM Microtubules; Nuclear protein.  
SQ SEQUENCE 846 AA; 98226 MW; 803048B05E5E105 CRC64;  
  
Query Match 2.0%; Score 7; DB 1; Length 846;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 88 LPPGSYI 94  
DB 608 LPPGSYI 614  
  
RESULT 38  
SP98\_YEAST STANDARD; PRT; 846 AA.  
AC P33540;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE SPINDLE POLE BODY COMPONENT SPC98.  
GN SPC98 OR YNL126W OR N122 OR N1879.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=5288C;  
RX MEDLINE=96109393; PubMed=8619318;  
RA Mallet L., Bussereau F., Jacquet M.;  
RT "A 43.5 kb segment of yeast chromosome XIV, which contains MFA2,  
RT MRP2, CAP/SPV2, NAW9, FKBI/PPRI/RBPI, MOM22 and CP11, predicts an  
RT adenosine deaminase gene and 14 new open reading frames.";  
RL Yeast 11:1195-1209(1995).  
RN [2]

OY 148 SYTLRQL 154  
DB 164 SYTLRQL 170  
  
RESULT 39  
CADB\_XENLA STANDARD; PRT; 905 AA.  
AC P33152;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE BLASTOMERE-CADHERIN PRECURSOR (B-CADHERIN).  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95151580; PubMed=7531482;  
RA Mueller H.-A.J., Kuehl M., Finemann S., Schneider S.,  
RA van der Poel S.Z., Hausen P., Wedlich D.;  
RT "Xenopus cadherins: the maternal pool comprises distinguishable  
RT members of the family";  
RL Mech. Dev. 47:213-223(1994).  
RN [2]  
RP SEQUENCE OF 459-905 FROM N.A.  
RX MEDLINE=92062581; PubMed=1840622;  
RA Herzberg F., Wildermuth V., Wedlich D.;  
RT "Expression of Xcad, a novel cadherin, during oogenesis and early  
RT development of Xenopus.";



```

DE METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)
OS Odontoglossum ringspot virus (Isolate Korean Cy) (ORSV-Cy)
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
OX NCBI_TaxID=138661;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96381046; PubMed=8789059;
RA Ikegami M., Isumura Y., Matsumoto Y., Chatani M., Inoue N.;
RT "The complete nucleotide sequence of odontoglossum ringspot virus (Cy-1 strain) genomic RNA."
RL Microbiol. Immunol. 39:995-1001(1995).
CC -1- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL RNA REPLICATION.
CC -1- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA CAPING AND AN RNA HELICASE.
CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS BETWEEN CODONS FOR LEU-1112 AND GLN-1114.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D13941; AAA21828.1;
DR EMBL: S83257; AAB49498.1;
DR InterPro: IPR000606;
DR InterPro: IPR001788;
DR InterPro: IPR002588;
DR Pfam: PF00978; RNA_dep_RNApol2; 1.
DR Pfam: PF01443; Viral_helicase; 1.
DR Pfam: PF01660; Vmethyltransf; 1.
KW transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT CHAIN 1 1612 METHYLTRANSFERASE/RNA POLYMERASE.
FT CHAIN 1 1112 METHYLTRANSFERASE/RNA HELICASE.
FT NP_BIND 826 833 ATP (POTENTIAL).
SQ SEQUENCE 1612 AA; 183252 MW; 3F803A2B9611EDFC C664;

```

```

Query Match 2.0%, Score 7; DB 1; Length 1612;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 227 GEDSEOF 233
DB 325 GEDSEOF 331

RESULT 42
ID RPO_ORSVS STANDARD; PRT; 1612 AA.
AC 084133;
DT 01-OCT-2000 (Rel. 40; Last sequence update)
DT 01-OCT-2000 (Rel. 40; Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS: METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS Odontoglossum ringspot virus (Isolate Singapore 1) (ORSV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
OX NCBI_TaxID=138662;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96257213; PubMed=8666266;
RA Chung C.G., Wong S.M., Mahant P.H., Loh C.S., Goh C.J., Kuo M.C.C.,
RA Chung M.C.M., Watanabe Y.;
RT "The complete sequence of a Singapore isolate of odontoglossum ringspot virus and comparison with other tobamoviruses."
RL Gene 171:155-161(1996).
CC -1- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL RNA REPLICATION.
CC -1- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA

```

```

CC CAPING AND AN RNA HELICASE.
CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS BETWEEN CODONS FOR LEU-1112 AND GLN-1114.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U34586; AAC55012.1;
DR InterPro: IPR000606;
DR InterPro: IPR001788;
DR InterPro: IPR002588;
DR Pfam: PF00978; RNA_dep_RNApol2; 1.
DR Pfam: PF01443; Viral_helicase; 1.
DR Pfam: PF01660; Vmethyltransf; 1.
KW transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT CHAIN 1 1612 RNA-DIRECTED RNA POLYMERASE.
FT CHAIN 1 1112 METHYLTRANSFERASE/RNA HELICASE.
FT NP_BIND 826 833 ATP (POTENTIAL).
SQ SEQUENCE 1612 AA; 183194 MW; 3B444670B0ACB189 C664;

```

```

Query Match 2.0%, Score 7; DB 1; Length 1612;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 227 GEDSEOF 233
DB 325 GEDSEOF 331

```

```

RESULT 43
ID MYSE_RAT STANDARD; PRT; 1940 AA.
AC P12847;
DT 01-OCT-1989 (Rel. 12; Created)
DT 01-OCT-1989 (Rel. 12; Last sequence update)
DT 15-JUL-1998 (Rel. 36; Last annotation update)
DE MYOSIN HEAVY CHAIN, FAST SKELETAL MUSCLE, EMBRYONIC.
GN MYH3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87060988; PubMed=3783701;
RA Strehler E.E., Strehler-Page M.-A., Perliard J.C., Periasamy M.,
RA Nadal-Ginard B.;
RT "Complete nucleotide and encoded amino acid sequence of a mammalian myosin heavy chain gene. Evidence against intron-dependent evolution of the rod."
RL J. Mol. Biol. 190:291-317(1986).
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
CC -1- DOMAIN: THE ROD-LIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES, CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- PTM: TWO CISTEINE RESIDUES IN THE SI DOMAIN ARE SELECTIVELY ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HWM). IT CAN LATER BE SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).
CC -1- MISCELLANEOUS: THERE ARE 10 OR MORE MYOSIN HEAVY CHAIN GENES IN THE RAT.

```

```

CC -----
CC -1- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X04267; CAA27817.1; -.
CC PTR: A24922; A24922.
CC HSSP: P13538; 2MS.
CC InterPro: IPR000048; -.
CC InterPro: IPR001609; -.
CC InterPro: IPR002928; -.
CC Pfam: PF00612; IO: 1.
CC Pfam: PF01576; Myosin_tail; 1.
CC Pfam: PF00063; myosin_head; 1.
CC PRINTS: PR00193; MYOSINHEAVY.
CC PROSITE: PSS0096; IO: 1.
CC Myosin: Muscle protein; Coiled coil; Thick filament; Actin-binding;
CC ATP-binding; Methylation; Alkylation; Multigene family.
CC K1M
CC Myosin: Muscle protein; Coiled coil; Thick filament; Actin-binding;
CC ATP-binding; Methylation; Alkylation; Multigene family.
CC FT DOMAIN 1 839 GLOBULAR HEAD (SI).
CC FT DOMAIN 840 1940 RODLIKE TAIL (S2 AND LMM DOMAINS).
CC FT NP_BIND 840 1933 COILED COIL (POTENTIAL).
CC FT DOMAIN 179 186 ATP (POTENTIAL).
CC FT DOMAIN 656 678 ACTIN-BINDING.
CC FT DOMAIN 758 772 ACTIN-BINDING.
CC FT MOD_RES 130 130 METHYLATION (TR1-) (POTENTIAL).
CC FT MOD_RES 696 696 ALKYLATION (SH-1).
CC FT MOD_RES 706 706 ALKYLATION (SH-2).
CC SEQUENCE 1940 AA; 223857 MW; B5D546A596E5A696 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 1940;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 ELAGIDE 221
Db 982 ELAGIDE 988

RESULT 44
YRS9_CABEL STANDARD; PRT; 2329 AA.
ID YRS9_CABEL
AC 009624;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOTHEICAL 254.3 KDA PROTEIN ZK945.9 IN CHROMOSOME II.
GN ZK945.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Wilkinson-Sprat J.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z48544; CAA84442.1; -.

```

```

DR WormPep: ZK945.9; CE01740.
DR InterPro: IPR000203; -.
DR InterPro: IPR001024; -.
DR Pfam: PF01825; GFS; 1.
DR Pfam: PF01477; PLAT; 1.
KW Hypothetical protein; Repeat; Transmembrane.
FT DOMAIN 22 115 SER/THR-RICH.
FT DOMAIN 250 389 GLY/SER-RICH.
FT TRANSMEM 557 577 POTENTIAL.
FT TRANSMEM 606 626 POTENTIAL.
FT TRANSMEM 1162 1182 POTENTIAL.
FT TRANSMEM 1290 1310 POTENTIAL.
FT TRANSMEM 1467 1487 POTENTIAL.
FT TRANSMEM 1498 1518 POTENTIAL.
FT TRANSMEM 1541 1561 POTENTIAL.
FT TRANSMEM 1602 1622 POTENTIAL.
FT TRANSMEM 1637 1657 POTENTIAL.
FT TRANSMEM 1717 1737 POTENTIAL.
FT TRANSMEM 1935 1955 POTENTIAL.
FT TRANSMEM 1990 2010 POTENTIAL.
FT TRANSMEM 2039 2059 POTENTIAL.
FT TRANSMEM 2088 2108 POTENTIAL.
FT TRANSMEM 2134 2154 POTENTIAL.
FT TRANSMEM 2189 2209 POTENTIAL.
SQ SEQUENCE 2329 AA; 254348 MW; 4D9D42306F0C85 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 2329;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 162 NSVLSL 168
Db 530 NSVLSL 536

RESULT 45
RS2_BACE STANDARD; PRT; 67 AA.
ID RS2_BACE
AC 09XBK3;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 30S RIBOSOMAL PROTEIN S2 (FRAGMENT).
DE RPSB.
OS Bacillus cereus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_Taxid=1396;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10987;
RX MEDLINE=99231848; PubMed=10217496;
RA Oksa O.A., Hegna I., Lindbeck T., Rishovd A.L., Kolsto A.B.;
RT "Genome organization is not conserved between Bacillus cereus and
RT Bacillus subtilis."
RL Microbiology 145:621-631(1999).
CC -1- SIMILARITY: BELONGS TO THE S2P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: A010135; CAB40591.1; -.
CC InterPro: IPR001865; -.
CC Pfam: PF00318; Ribosomal_S2; 1.
CC PRINTS: PR00395; RIBOSOMALS2.
DR PROSITE: PSS00962; RIBOSOMAL_S2_1; PARTIAL.
DR PROSITE: PSS00963; RIBOSOMAL_S2_2; PARTIAL.

```

KM Ribosomal protein.  
 FT NON\_TER 1  
 SQ SEQUENCE 67 AA: 7303 MW: 04182C7BD47D89FD CRC64:

Query Match  
 Best Local Similarity 100.0%; Score 6; DB 1; Length 67;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 ILEAKQ 285  
 DB 55 ILEAKQ 60

RESULT 46  
 NUAM\_ARTSA STANDARD; PRT: 71 AA.

AC P19046:  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 4 (EC 1.6.5.3) (FRAGMENT).  
 GN NDA.  
 OS Artemia salina (Brine shrimp).  
 OG Mitochondrion.  
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Branchiopoda; Anostraca;  
 OC Artemiidae; Artemia.  
 OX NCBI\_TaxID=85349;

RA MEDLINE=88289417; PubMed=3135541;  
 RA Batuecas B., Garesse R., Calleja M., Valverde J. R., Marco R.;  
 RT "Genome organization of Artemia mitochondrial DNA."  
 RL Nucleic Acids Res. 16:6515-6529(1988).  
 CC -I- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.  
 CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: X07662; CA30509.1; -.  
 CC PIR: S01211; S01211.  
 DR InterPro: IPR001750; -.  
 DR Pfam: PF00361; oxidored\_4t.1.  
 KM Oxidoreductase; NAD; Ubiquinone; Mitochondrion.  
 FT NON\_TER 1  
 FT NON\_TER 71  
 SQ SEQUENCE 71 AA: 8524 MW: D7E735D6ADB7A7 CRC64:

Query Match  
 Best Local Similarity 100.0%; Score 6; DB 1; Length 71;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 LSSDPL 170  
 DB 54 LSSDPL 59

RESULT 47  
 IGFL\_SUNMU STANDARD; PRT: 81 AA.  
 AC Q28933:  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE INSULIN-LIKE GROWTH FACTOR I PRECURSOR (IGF-I) (SOMATOMEDIN)  
 DE (FRAGMENT).  
 GN IGFL.

OS Suncus murinus (house shrew) (musk shrew).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Insectivora; Soricidae; Suncus.  
 OX NCBI\_TaxID=9378;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BAN, AND NAG; TISSUE=Liver;  
 RA Ishikawa A.;

RT "Partial sequence of a IGF-I cDNA in the musk shrew, Suncus murinus."  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 CC -I- FUNCTION: THE INSULIN-LIKE GROWTH FACTORS, ISOLATED FROM PLASMA,  
 CC ARE STRUCTURALLY AND FUNCTIONALLY RELATED TO INSULIN BUT HAVE A  
 CC MUCH HIGHER GROWTH-PROMOTING ACTIVITY.  
 CC -I- SUBCELLULAR LOCATION: SECRETED.  
 CC -I- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: DA3957; BAA07897.1; -.  
 DR HSSP: P05019; 3GF1.  
 DR InterPro: IPR000739; -.  
 DR Pfam: PF00049; Insulin; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 KW Insulin family; Growth factor; Plasma.  
 FT NON\_TER 1  
 FT PROPEP 1  
 FT CHAIN 4  
 FT DOMAIN 5  
 FT DOMAIN 5  
 FT DOMAIN 34  
 FT DOMAIN 46  
 FT DOMAIN 67  
 FT PROPEP 75  
 FT DISULFID 10  
 FT DISULFID 22  
 FT DISULFID 51  
 FT NON\_TER 81  
 SQ SEQUENCE 81 AA: 8869 MW: AC2C40972D05E3C4 CRC64:

FT 1 BY SIMILARITY.  
 FT 4 INSULIN-LIKE GROWTH FACTOR I.  
 FT 74 B.  
 FT 75 C.  
 FT 76 A.  
 FT 77 D.  
 FT 78 E PEPTIDE.  
 FT 79 BY SIMILARITY.  
 FT 80 BY SIMILARITY.  
 FT 81 BY SIMILARITY.

Query Match  
 Best Local Similarity 100.0%; Score 6; DB 1; Length 81;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 240 LVDAIQ 245  
 DB 14 LVDAIQ 19

RESULT 48  
 YN3\_PSEST STANDARD; PRT: 82 AA.  
 AC P95549:  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE HYPOTHETICAL 8.5 KDA PROTEIN IN N1R0 3/REGION (ORF82).  
 OS Pseudomonas stutzeri (Pseudomonas perfectmarina).  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=316;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 14405 / ZOBELL;  
 RX MEDLINE=97107629; PubMed=8950369;  
 RA Glockner A.B., Zumft W.G.;

RT "Sequence analysis of an internal 9.72-kb segment from the 30-kb  
 denitrification gene cluster of Pseudomonas stutzeri."

```

RL Blochim. Biophys. Acta 1277:6-12(1996).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z73914; CAA98150.1; -.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 22 42 POTENTIAL.
FT TRANSMEM 62 82 POTENTIAL.
SQ SEQUENCE 82 AA; 8523 MW; BFA5B5D3DAV34505 CRC64;

Query Match
Best Local Similarity 1.7%; Score 6; DB 1; Length 82;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 ASKTLV 241
   |||||
DB 3 ASKTLV 8

RESULT 49
PE60_HUMAN STANDARD; PRT; 86 AA.
ID PE60_HUMAN
AC 060575;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PEPTIDE PEC-60 HOMOLOG PRECURSOR.
GN SPINK4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Hu G.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PROTEASE INHIBITORS KAZAL FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF048700; AAC05124.1; -.
DR HSSP: P37109; 1PCE.
DR InterPro: IPR002350; -.
DR Pfam: PFO0050; Kazal.1.
DR PROSITE: PS00282; KAZAL.1.
KW Serine protease inhibitor; Signal.
FT SIGNAL 1 26 BY SIMILARITY.
FT CHAIN 27 86 PEPTIDE PEC-60 HOMOLOG.
FT DISULFD 37 68 BY SIMILARITY.
FT DISULFD 46 65 BY SIMILARITY.
FT DISULFD 54 86 BY SIMILARITY.
FT ACT_SITE 48 49 REACTIVE BOND (BY SIMILARITY).
SQ SEQUENCE 86 AA; 9434 MW; 374CDB30AC9AF52 CRC64;

Query Match
Best Local Similarity 1.7%; Score 6; DB 1; Length 86;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 IATALA 320
   |||||
DB 8 IATALA 13

RESULT 50
VL02_VACCC STANDARD; PRT; 87 AA.
ID VL02_VACCC
AC P20843;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE PROTEIN L2.
GN L2R.
OS Vaccinia virus (strain Copenhagen).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10249;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021027; PubMed=2219722;
RA Goebel S.J., Johnson G.P., Perkins M.E., Davis S.W., Winslow J.P.,
RA Paoletti E.;
RT "The complete DNA sequence of vaccinia virus."
RL Virology 179:247-266(1990).
RN [2]
RP COMPLETE GENOME.
RA Goebel S.J., Johnson G.P., Perkins M.E., Davis S.W., Winslow J.P.,
RA Paoletti E.;
RL Virology 179:517-563(1990).
CC -1- SIMILARITY: TO FOMLOPOX PROTEIN FP3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M35027; AAA48077.1; -.
DR PIR: I42512; I42512.
SQ SEQUENCE 87 AA; 10225 MW; C3BA27F8E59C448E CRC64;

Query Match
Best Local Similarity 1.7%; Score 6; DB 1; Length 87;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LLAIVA 12
   |||||
DB 69 LLAIVA 74

```

Search completed: August 17, 2001, 16:46:51  
Job time: 118 sec

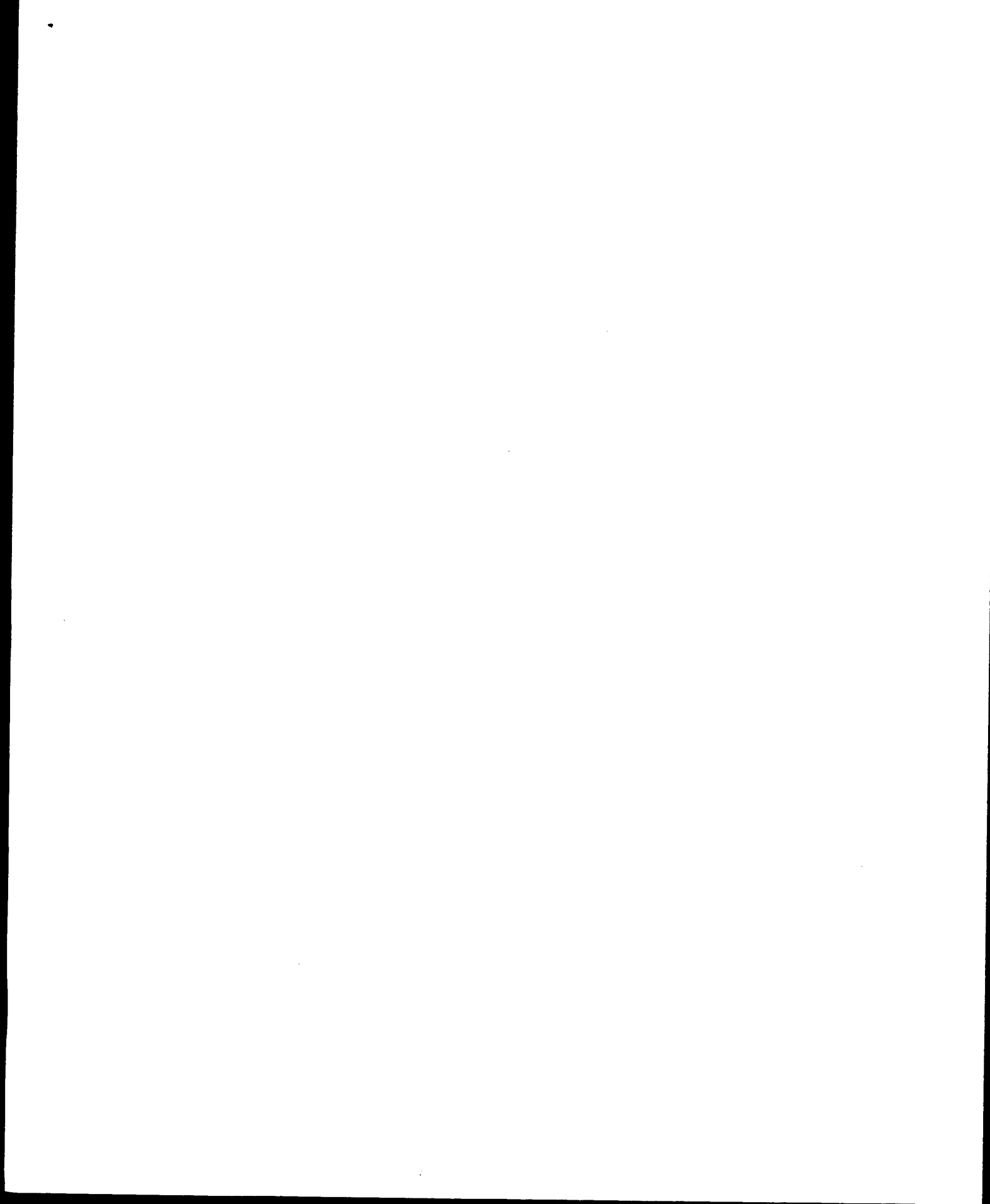


Mon Aug 20 10:21:53 2001

us-09-284-320-6.oli5.rsp

---

Page 31



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 17, 2001, 16:43:58 ; Search time 16.14 Seconds  
(without alignments)  
1651.865 Million cell updates/sec

Title: US-09-284-320-6

Perfect score: 350

Sequence: 1 MAVPVLLALVAVGLGNFES.....MDPGYDIIYRMTNQKLRMD 350

Scoring table: OLIGO  
Gapop 60.0 , Gapept 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 5

Total number of hits satisfying chosen parameters: 23256

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	264	75.4	330	2	T08667
2	9	2.6	343	2	T44797
3	8	2.3	55	2	T36132
4	8	2.3	156	2	B84227
5	8	2.3	223	2	E86728
6	8	2.3	324	2	T3876
7	8	2.3	468	2	F86256
8	8	2.3	556	2	T10568
9	8	2.3	1413	2	D84481
10	8	2.3	1501	2	C84512
11	8	2.3	1715	2	C40228
12	8	2.3	5149	2	F83345
13	7	2.0	29	2	CS4037
14	7	2.0	73	2	S42862
15	7	2.0	74	2	T26655
16	7	2.0	109	2	T44129
17	7	2.0	112	2	A45069
18	7	2.0	119	2	G65203
19	7	2.0	120	2	S21565
20	7	2.0	137	2	B64208
21	7	2.0	137	2	S14025
22	7	2.0	148	1	D65349
23	7	2.0	148	1	D65349
24	7	2.0	151	2	C83091
25	7	2.0	151	2	C83091
26	7	2.0	163	1	CFYCA
27	7	2.0	168	2	G75525
28	7	2.0	179	2	C69467
29	7	2.0	184	2	C72495

30	7	2.0	195	2	S28739
31	7	2.0	196	2	T29016
32	7	2.0	203	2	S57836
33	7	2.0	203	2	T45907
34	7	2.0	213	2	F70476
35	7	2.0	213	2	T18061
36	7	2.0	224	2	E72401
37	7	2.0	228	2	A83413
38	7	2.0	238	1	GOORPE
39	7	2.0	239	2	B84249
40	7	2.0	240	2	T38152
41	7	2.0	254	2	E71305
42	7	2.0	260	2	D75404
43	7	2.0	266	2	T39301
44	7	2.0	267	2	G69791
45	7	2.0	274	2	B83215
46	7	2.0	274	2	E83909
47	7	2.0	279	2	E75188
48	7	2.0	281	2	T00880
49	7	2.0	291	2	B70928
50	7	2.0	299	2	S76631
51	7	2.0	310	2	C70871
52	7	2.0	321	2	B72384
53	7	2.0	326	2	H83187
54	7	2.0	327	2	S30405
55	7	2.0	336	2	F75320
56	7	2.0	337	2	D64896
57	7	2.0	337	2	B85734
58	7	2.0	338	2	H72679
59	7	2.0	349	2	A40551
60	7	2.0	356	2	E64048
61	7	2.0	356	2	T10888
62	7	2.0	363	2	T41209
63	7	2.0	363	2	T42371
64	7	2.0	364	2	T30084
65	7	2.0	364	2	T39235
66	7	2.0	385	2	T21763
67	7	2.0	386	2	T05691
68	7	2.0	388	2	A82903
69	7	2.0	394	2	C36942
70	7	2.0	394	2	G71155
71	7	2.0	395	2	T81358
72	7	2.0	396	2	T45033
73	7	2.0	398	2	A59371
74	7	2.0	400	2	S74909
75	7	2.0	401	2	E82951
76	7	2.0	401	2	A69368
77	7	2.0	404	2	A45670
78	7	2.0	414	2	T44711
79	7	2.0	416	2	A70779
80	7	2.0	426	2	S58132
81	7	2.0	436	2	E59237
82	7	2.0	439	2	S20489
83	7	2.0	450	2	S57837
84	7	2.0	452	2	S67569
85	7	2.0	506	2	C81190
86	7	2.0	506	2	A81914
87	7	2.0	515	2	T36714
88	7	2.0	540	2	T36409
89	7	2.0	555	2	E70409
90	7	2.0	571	2	A72555
91	7	2.0	576	2	A71497
92	7	2.0	612	2	T11628
93	7	2.0	612	2	E82153
94	7	2.0	638	2	B28176
95	7	2.0	646	2	T19206
96	7	2.0	648	2	T23864
97	7	2.0	729	2	T04259
98	7	2.0	744	2	H71643
99	7	2.0	770	2	T00204
100	7	2.0	770	2	T00203
101	7	2.0	788	2	S67595
102	7	2.0	793	1	S60735

hypothetical prote  
hypothetical prote  
lymphoid-specific  
hypothetical prote  
conserved hypothet  
hypothetical prote  
conserved hypothet  
molybdenum transpo  
tRNA nucleotidyltr  
hypothetical prote  
conserved hypothet  
phosphatidylglycer  
probable O-methyl  
hypothetical prote  
probable permease  
oligopeptide ABC t  
morphine 6-dehydro  
hypothetical prote  
probable dach prot  
hypothetical prote  
hypothetical prote  
bacterioferritin c  
hypothetical prote  
hypothetical prote  
WD-repeat family p  
probable ABC-type  
hypothetical prote  
hypothetical prote  
connective tissue  
nucleotide-binding  
iron utilization p  
mannose-1-phosphat  
probable mannose-1  
hypothetical prote  
probable translati  
hypothetical prote  
multidrug-resista  
conserved hypothet  
hypothetical prote  
hypothetical prote  
transmembrane tran  
hypothetical prote  
Ig V-region-like B  
transposase slt206  
probable MFS trans  
hypothetical prote  
2',3'-cyclic-nucle  
beta-ketocacyl-ACP  
probable kasp prot  
SL1 protein precu  
probable glutamate  
S-locus-specific g  
lymphoid-specific  
hypothetical prote  
conserved hypothet  
probable integral  
probable ABC-type  
carbon starvation  
probable cytochrom  
probable DNA mima  
WPI protein - fls  
PTS system, fructo  
somatotropin recep  
hypothetical prote  
hypothetical prote  
probable beta-gala  
cell division prot  
LDL receptor relat  
LDL receptor-relat  
hypothetical prote  
splicing factor SF

103	7	2.0	808	2	E64492	176	6	1.7	129	2	B84595	hypothetical prote
104	7	2.0	816	2	D66544	177	6	1.7	130	2	UC2280	ribosomal protein
105	7	2.0	846	2	S59262	178	6	1.7	130	2	UC2280	hypothetical 14.0k
106	7	2.0	847	2	T12979	179	6	1.7	131	1	S12972	amicyanin - Paraco
107	7	2.0	863	2	F84504	180	6	1.7	131	2	S75262	sensory transduct
108	7	2.0	905	2	S43064	181	6	1.7	133	2	G83160	conserved hypotet
109	7	2.0	974	2	A72012	182	6	1.7	133	2	A40912	insulin-like growt
110	7	2.0	974	2	B86613	183	6	1.7	133	2	A75202	hypothetical prote
111	7	2.0	1112	2	S49432	184	6	1.7	133	2	A75202	natriuretic peptid
112	7	2.0	1121	2	A82809	185	6	1.7	134	1	AMHDB	gastrin-releasing
113	7	2.0	1140	2	A34728	186	6	1.7	134	2	147010	PHL E1F1 - human
114	7	2.0	1156	2	D84639	187	6	1.7	135	2	H70516	hypothetical prote
115	7	2.0	1171	2	P83110	188	6	1.7	136	1	B64414	hypothetical prote
116	7	2.0	1304	2	T14073	189	6	1.7	136	2	T10551	hypothetical prote
117	7	2.0	1762	2	T03222	190	6	1.7	136	2	S44623	C50c3.4 protein -
118	7	2.0	1940	1	A24922	191	6	1.7	136	2	T49582	hypothetical prote
119	7	2.0	2076	2	T28915	192	6	1.7	137	1	TCGP1	insulin-like growt
120	7	2.0	2100	2	T03223	193	6	1.7	137	2	S43885	3-Isopropylmalate
121	7	2.0	2329	2	T28125	194	6	1.7	137	2	A56552	insulin-like growt
122	7	2.0	2340	2	T48310	195	6	1.7	137	2	B72736	hypothetical prote
123	6	1.7	30	2	S08565	196	6	1.7	138	2	A26182	gastrin-releasing
124	6	1.7	52	2	G82639	197	6	1.7	138	2	S22878	insulin-like growt
125	6	1.7	67	2	G39098	198	6	1.7	139	2	A70881	insulin-like growt
126	6	1.7	68	2	H59098	199	6	1.7	139	2	T30863	hypothetical prote
127	6	1.7	71	2	S01211	200	6	1.7	139	2	F84024	stb protein - Sal
128	6	1.7	76	2	S72752	201	6	1.7	142	2	F84975	arsenate reductase
129	6	1.7	79	2	S75193	202	6	1.7	142	2	A39418	50S ribosomal prot
130	6	1.7	80	2	C49050	203	6	1.7	143	2	S13238	DNA-directed RNA p
131	6	1.7	82	2	T51104	204	6	1.7	143	2	T23735	tail fiber assembl
132	6	1.7	82	2	B83740	205	6	1.7	144	2	F28585	hypothetical prote
133	6	1.7	85	2	S77937	206	6	1.7	145	2	S26443	protein D - Escher
134	6	1.7	87	2	I42512	207	6	1.7	146	2	H84358	hypothetical prote
135	6	1.7	87	2	J01683	208	6	1.7	148	1	B26182	gastrin-releasing
136	6	1.7	88	2	T12130	209	6	1.7	148	2	S46514	putridolone-a pre
137	6	1.7	88	2	C82627	210	6	1.7	149	2	T38733	purindoline-b pre
138	6	1.7	89	2	D75271	211	6	1.7	150	2	A81845	hypothetical prote
139	6	1.7	91	2	A64760	212	6	1.7	150	2	G81096	50S ribosomal prot
140	6	1.7	91	2	T06008	213	6	1.7	152	2	T32784	50S ribosomal prot
141	6	1.7	92	1	Q0VZF3	214	6	1.7	152	2	S32567	hypothetical prote
142	6	1.7	92	1	G86351	215	6	1.7	153	1	IGHU1	insulin-like growt
143	6	1.7	94	2	F31844	216	6	1.7	153	1	IGHU1	insulin-like growt
144	6	1.7	96	1	ZRBPL	217	6	1.7	153	1	IGHU1	insulin-like growt
145	6	1.7	96	1	A85847	218	6	1.7	153	1	B27804	insulin-like growt
146	6	1.7	97	2	C75415	219	6	1.7	153	2	S12825	insulin-like growt
147	6	1.7	101	2	T45325	220	6	1.7	153	2	A41399	insulin-like growt
148	6	1.7	104	2	S04103	221	6	1.7	154	2	JC2483	insulin-like growt
149	6	1.7	108	2	C72451	222	6	1.7	154	2	A33390	insulin-like growt
150	6	1.7	109	2	S41018	223	6	1.7	155	2	A64087	insulin-like growt
151	6	1.7	110	2	PQ0494	224	6	1.7	155	2	S74527	xanthine phosphor
152	6	1.7	110	2	C71076	225	6	1.7	156	2	H71023	hypothetical prote
153	6	1.7	111	2	D82986	226	6	1.7	157	2	T49394	hypothetical prote
154	6	1.7	112	1	OOKMTB	227	6	1.7	158	2	S53572	probable membrane
155	6	1.7	112	2	B47033	228	6	1.7	158	2	H82937	hypothetical prote
156	6	1.7	114	2	B71221	229	6	1.7	159	2	S68978	Na+/Ca2+,K+-exchan
157	6	1.7	114	2	B69966	230	6	1.7	159	2	A26859	insulin-like growt
158	6	1.7	115	2	H83430	231	6	1.7	159	2	T10837	probable resistanc
159	6	1.7	116	2	A83605	232	6	1.7	161	2	S14576	globin E precursor
160	6	1.7	117	2	B86601	233	6	1.7	161	2	D39441	sporIGA homolog
161	6	1.7	117	2	H72022	234	6	1.7	162	1	F72631	hypothetical prote
162	6	1.7	119	2	T47169	235	6	1.7	162	1	C7FC43	R-phycoeryanin II a
163	6	1.7	120	2	PC4251	236	6	1.7	163	2	T11215	protein F12A21.27
164	6	1.7	121	2	S59102	237	6	1.7	166	2	S36294	conserved hypotet
165	6	1.7	122	2	PN0622	238	6	1.7	166	2	B56986	conserved hypotet
166	6	1.7	123	2	JC1202	239	6	1.7	168	2	F69072	T-cell receptor ga
167	6	1.7	123	2	S29304	240	6	1.7	169	2	E64462	conserved hypotet
168	6	1.7	123	2	G82357	241	6	1.7	169	2	C96551	hypothetical prote
169	6	1.7	126	2	S66485	242	6	1.7	169	2	T35376	hypothetical prote
170	6	1.7	126	2	H69357	243	6	1.7	171	2	A72333	hypothetical prote
171	6	1.7	127	2	B40912	244	6	1.7	173	2	C86503	acetyltransferase [i
172	6	1.7	127	2	A25540	245	6	1.7	174	2	T07736	probable chaperoni
173	6	1.7	127	2	A64243	246	6	1.7	174	2	A70828	hypothetical prote
174	6	1.7	128	2	E70466	247	6	1.7	176	1	S74642	hypothetical prote
175	6	1.7	128	2	S25103	248	6	1.7	176	1	S74642	hypothetical prote

249	6	1.7	176	2	D81209	322	6	1.7	213	2	T46078	hypothetical prote
250	6	1.7	176	2	H72201	323	6	1.7	213	2	S74055	hypothetical prote
251	6	1.7	177	2	T07286	324	6	1.7	214	2	B46244	insulin-like growt
252	6	1.7	177	2	T21952	325	6	1.7	214	2	T10737	extensin-like cell
253	6	1.7	177	2	F72715	326	6	1.7	214	2	T09854	proline-rich cell
254	6	1.7	178	2	D75390	327	6	1.7	215	2	E72736	probable DNA polym
255	6	1.7	180	2	T41322	328	6	1.7	215	2	B83429	conserved hypotbet
256	6	1.7	180	2	A82493	329	6	1.7	216	2	F82013	capsule polysaccha
257	6	1.7	181	2	A27804	330	6	1.7	216	2	H81241	capsule polysaccha
258	6	1.7	181	2	G70111	331	6	1.7	217	2	D96658	hypothetical prote
259	6	1.7	181	2	H72853	332	6	1.7	217	2	A84810	probable guanilate
260	6	1.7	182	2	F84427	333	6	1.7	218	2	T01076	transcription fact
261	6	1.7	183	2	A72499	334	6	1.7	218	2	E86750	hypothetical prote
262	6	1.7	184	1	H64845	335	6	1.7	219	1	A29128	myelin P0 protein
263	6	1.7	184	2	A85665	336	6	1.7	219	2	C65074	hypothetical prote
264	6	1.7	184	2	T42292	337	6	1.7	219	2	G85945	probable oxidoredu
265	6	1.7	184	2	D84008	338	6	1.7	220	2	T28639	Y41L protein - Rht
266	6	1.7	185	2	E84957	339	6	1.7	220	2	C4383	hypothetical prote
267	6	1.7	185	2	T28898	340	6	1.7	221	2	C46482	T-cell surface gly
268	6	1.7	185	2	H83296	341	6	1.7	222	2	E70374	alkyl hydroperoxid
269	6	1.7	185	2	F75273	342	6	1.7	222	2	E71844	hypothetical prote
270	6	1.7	189	1	F69951	343	6	1.7	222	2	G64670	hypothetical prote
271	6	1.7	190	2	T37428	344	6	1.7	222	2	A82958	hypothetical prote
272	6	1.7	191	2	S76398	345	6	1.7	222	2	C84886	probable PREG1-lik
273	6	1.7	191	2	F75340	346	6	1.7	222	2	T03565	ribosomal protein
274	6	1.7	192	2	T39464	347	6	1.7	224	1	B70106	conserved hypotbet
275	6	1.7	193	2	H84051	348	6	1.7	224	2	D70725	hypothetical prote
276	6	1.7	194	2	S64739	349	6	1.7	224	2	E64542	hypothetical prote
277	6	1.7	194	2	T47200	350	6	1.7	224	2	T34686	probable integral
278	6	1.7	194	2	T42521	351	6	1.7	224	2	J01512	hypothetical 26.2K
279	6	1.7	194	2	T01780	352	6	1.7	225	2	T02660	germin-like protei
280	6	1.7	194	2	D72242	353	6	1.7	225	2	B84653	TINW-like AP2 doma
281	6	1.7	194	2	G75328	354	6	1.7	225	2	T28320	ORF MSV159 hypotbe
282	6	1.7	194	2	G72517	355	6	1.7	226	2	A86268	hypothetical prote
283	6	1.7	195	1	S74599	356	6	1.7	226	2	D69780	hypothetical prote
284	6	1.7	195	1	IGHU1B	357	6	1.7	227	2	C31390	pHb protein - Ser
285	6	1.7	195	2	T28588	358	6	1.7	227	2	S77253	precocarin decarbox
286	6	1.7	195	2	D36853	359	6	1.7	227	2	B69503	conserved hypotbet
287	6	1.7	195	2	H72169	360	6	1.7	228	2	G68682	hypothetical prote
288	6	1.7	195	2	T36935	361	6	1.7	228	2	D81917	probable transcrip
289	6	1.7	195	2	A75295	362	6	1.7	228	2	S41760	heat shock protein
290	6	1.7	197	2	S55621	363	6	1.7	229	2	F75482	transcription regu
291	6	1.7	197	2	D70017	364	6	1.7	230	1	Q012V5	VII protein - Maed
292	6	1.7	197	2	A86731	365	6	1.7	230	2	E70927	probable trmb prot
293	6	1.7	197	2	B83561	366	6	1.7	230	2	F65042	probable membrane
294	6	1.7	197	2	T05252	367	6	1.7	230	2	H75343	SUA5-related prote
295	6	1.7	198	2	D71727	368	6	1.7	230	2	T47866	regulatory protein
296	6	1.7	200	2	G81304	369	6	1.7	231	2	B81424	probable amino aci
297	6	1.7	202	1	TYMSHS	370	6	1.7	231	2	F84233	competence-damage
298	6	1.7	203	2	S20586	371	6	1.7	231	2	B70795	probable cobQ2 pro
299	6	1.7	203	2	T21310	372	6	1.7	232	2	T05881	gene forever young
300	6	1.7	204	2	D81793	373	6	1.7	232	2	H85322	forever young gene
301	6	1.7	204	2	B81216	374	6	1.7	233	2	T12426	germin-like protei
302	6	1.7	205	2	T14310	375	6	1.7	233	2	I65197	hoxl.3 protein - r
303	6	1.7	205	2	C86862	376	6	1.7	233	2	F72671	hypothetical prote
304	6	1.7	205	2	G81909	377	6	1.7	234	2	H82730	ribonuclease HII X
305	6	1.7	206	2	B81105	378	6	1.7	234	2	S33732	myosin alpha heavy
306	6	1.7	206	2	H72649	379	6	1.7	234	2	P83191	conserved hypotbet
307	6	1.7	207	2	C70029	380	6	1.7	234	2	D85964	hypothetical prote
308	6	1.7	208	2	S15575	381	6	1.7	234	2	S22360	hypothetical prote
309	6	1.7	210	1	TPHUC	382	6	1.7	235	2	E96811	hypothetical prote
310	6	1.7	210	2	E46482	383	6	1.7	236	2	B69335	succinate dehydrog
311	6	1.7	210	2	S25657	384	6	1.7	236	2	JH0400	neurotrophin-4 pre
312	6	1.7	211	1	TPRATC	385	6	1.7	236	2	I57681	potassium channel
313	6	1.7	211	2	S65049	386	6	1.7	236	2	T48518	transcription fact
314	6	1.7	211	2	A29994	387	6	1.7	236	2	E71965	hypothetical prote
315	6	1.7	212	1	G64331	388	6	1.7	236	2	T04929	hypothetical prote
316	6	1.7	212	2	S65525	389	6	1.7	237	2	S40284	rad3 protein - mo
317	6	1.7	212	2	C64583	390	6	1.7	237	2	D64000	hypothetical prote
318	6	1.7	212	2	F71928	391	6	1.7	237	2	A65066	hypothetical prote
319	6	1.7	212	2	T48580	392	6	1.7	237	2	D85935	probable transport
320	6	1.7	213	2	G81836	393	6	1.7	237	2	T35855	probable ABC trans
321	6	1.7	213	2	B81128	394	6	1.7	237	2	B81298	hypothetical prote

395	1.7	237	2	B84377	hypothetical prote
396	1.7	238	2	UX0320	rhodanese-like pro
397	1.7	238	2	D75513	L-serine dehydrata
398	1.7	238	2	H70866	hypothetical prote
399	1.7	239	2	T20255	hypothetical prote
400	1.7	239	2	A81437	probable integral
401	1.7	239	2	T23147	hypothetical prote
402	1.7	240	2	S61110	hypothetical prote
403	1.7	240	2	H64373	hypothetical prote
404	1.7	241	1	S56541	chaperone protein
405	1.7	241	1	B69885	transcription regu
406	1.7	242	1	F86129	hypothetical prote
407	1.7	242	1	H71829	hypothetical prote
408	1.7	242	2	G64687	hypothetical prote
409	1.7	243	2	T16229	conserved hypotet
410	1.7	244	2	A34910	hypothetical prote
411	1.7	244	2	A46721	mast cell proteina
412	1.7	244	2	A46721	chymase (EC 3.4.21
413	1.7	244	2	A41021	NAD--asparagine A
414	1.7	244	2	S76413	hypothetical prote
415	1.7	244	2	S76413	hypothetical prote
416	1.7	244	2	T26704	hypothetical prote
417	1.7	245	2	T30127	hypothetical prote
418	1.7	245	2	E86536	hypothetical prote
419	1.7	245	2	G72086	hypothetical prote
420	1.7	245	2	T34978	hypothetical prote
421	1.7	246	2	T37169	ABC transporter At
422	1.7	246	2	D46482	hypothetical prote
423	1.7	246	2	T01073	T-cell surface gly
424	1.7	246	2	A69659	T cell surface gly
425	1.7	246	2	H72497	ribosomal protein
426	1.7	246	2	S42479	probable NH(3)-dep
427	1.7	246	2	D85948	CDM protein - huma
428	1.7	246	2	A65077	probable actin y99
429	1.7	246	2	D70724	hypothetical 26.6k
430	1.7	246	2	C83694	hypothetical prote
431	1.7	247	1	JC6540	hypothetical prote
432	1.7	247	2	JU0393	placenta specific-
433	1.7	247	2	JC5032	Karasurin - Mongol
434	1.7	247	2	A82421	Karasurin-B - Tric
435	1.7	248	2	S30194	arginine ABC trans
436	1.7	249	2	S75671	ribosomal protein
437	1.7	249	2	G84224	hypothetical prote
438	1.7	249	2	T04939	hypothetical prote
439	1.7	250	2	T20257	hypothetical prote
440	1.7	251	2	T11549	hypothetical prote
441	1.7	251	2	B86632	ATP binding protei
442	1.7	251	2	A84367	hypothetical prote
443	1.7	252	2	T33679	hypothetical prote
444	1.7	252	2	D82990	hypothetical prote
445	1.7	253	2	A70185	conserved hypotet
446	1.7	253	2	A70182	hepatocyte growth
447	1.7	253	2	E85587	phosphoglycerate m
448	1.7	253	2	F64815	hypothetical prote
449	1.7	254	2	D82963	yhhp protein - Esc
450	1.7	254	2	A44104	conserved hypotet
451	1.7	255	2	G69605	cholesterol Salph
452	1.7	255	2	S31096	hypothetical prote
453	1.7	256	2	T40964	proline-rich prote
454	1.7	256	2	S41185	probable hydroxyc
455	1.7	258	2	T05534	gene 38 protein -
456	1.7	259	2	I64916	protein-methionine
457	1.7	259	2	F81912	streptomycin 3'-a
458	1.7	260	2	T36343	hypothetical prote
459	1.7	261	1	SNRTG9	probable ABC-type
460	1.7	261	2	S38530	multicatalytic end
461	1.7	261	2	D81426	multicatalytic end
462	1.7	261	2	D84061	para family protei
463	1.7	262	2	F82959	ferrichrome ABC tr
464	1.7	262	2	JC6037	permease of ABC z
465	1.7	262	2	F85578	repressor protein
466	1.7	264	2	T09324	molybdate uptake r
467	1.7	264	2	F71466	capsid protein - h
					hypothetical prote
468	1.7	264	2	S63214	hypothetical prote
469	1.7	265	1	OTW3M	cytochrome-c oxida
470	1.7	265	1	OTW3M	cytochrome-c oxida
471	1.7	265	2	S20801	cytochrome-c oxida
472	1.7	265	2	T01578	probable membrane
473	1.7	266	2	F83127	probable short-cha
474	1.7	266	2	G83178	conserved hypotet
475	1.7	267	2	B75219	hypothetical prote
476	1.7	268	2	C83442	hypothetical prote
477	1.7	269	1	C69055	hypothetical prote
478	1.7	269	2	B23957	conserved hypotet
479	1.7	269	2	G75148	streptomycin/spect
480	1.7	269	2	G75148	hypothetical prote
481	1.7	270	1	WJH01C	hypothetical prote
482	1.7	270	1	WJH01C	homeotic protein H
483	1.7	270	2	C86674	hypothetical prote
484	1.7	270	2	C83167	hypothetical prote
485	1.7	271	2	T11704	probable transcrip
486	1.7	271	2	T11704	NADH dehydrogenase
487	1.7	271	2	S76871	hypothetical prote
488	1.7	272	2	T44929	hypothetical prote
489	1.7	272	2	F75412	moet protein (limo
490	1.7	272	2	S49912	spermidine/putresc
491	1.7	272	2	T20991	CATs protein - yea
492	1.7	273	2	B84987	hypothetical prote
493	1.7	273	2	B64980	hypothetical prote
494	1.7	274	1	C69362	shikimate 5-dehydr
495	1.7	275	1	R5K12	cofactors biosynth
496	1.7	275	2	A96552	conserved hypotet
497	1.7	275	2	A96552	ribosomal protein
498	1.7	276	2	A83063	unknown protein, 9
499	1.7	277	1	QRECS1	probable methyl tr
500	1.7	277	2	B85885	sulfate/hioulsulfat
501	1.7	277	2	B64548	hypothetical prote
502	1.7	277	2	B84950	conserved hypotet
503	1.7	277	2	B85701	heam protein (limo
504	1.7	277	2	G81134	probable protoporp
505	1.7	277	2	E81885	phosphoserine phos
506	1.7	278	2	T83570	probable phosphose
507	1.7	278	2	H72426	heam protein (EC 1
508	1.7	279	2	T08647	hypothetical prote
509	1.7	281	2	T25587	hypothetical prote
510	1.7	281	2	C72712	hypothetical prote
511	1.7	282	2	F69277	probable methionin
512	1.7	282	2	F70704	branched-chain am
513	1.7	283	2	S10773	hypothetical prote
514	1.7	283	2	E84070	2-hydroxymuconic s
515	1.7	283	2	E69419	spore coat polysac
516	1.7	283	2	S13383	phosphate ABC tran
517	1.7	283	2	B81288	hydroxyproline-ric
518	1.7	283	2	T01454	hypothetical prote
519	1.7	285	1	A37757	hypothetical prote
520	1.7	285	2	A37757	site-specific DNA-
521	1.7	285	2	S58240	acetylactate decar
522	1.7	285	2	C64750	hypothetical prote
523	1.7	285	2	S52681	probable transcrip
524	1.7	286	2	H86664	ribosomal protein
525	1.7	286	2	S71225	outer membrane lip
526	1.7	286	2	T18957	xyloglucan endo-1,
527	1.7	286	2	A82109	hypothetical prote
528	1.7	287	2	E75014	heam protein VC217
529	1.7	287	2	F72613	hypothetical prote
530	1.7	288	2	F85907	hypothetical prote
531	1.7	288	2	F72219	hypothetical prote
532	1.7	289	2	F65039	conserved hypotet
533	1.7	289	2	JC5606	hypothetical prote
534	1.7	291	2	S26028	Karasurin C - Tric
535	1.7	291	2	H71467	NADH dehydrogenase
536	1.7	291	2	H81729	succinate--CoA lig
537	1.7	291	2	A25686	succinyl-CoA synth
538	1.7	292	2	G71300	ampr protein - Bnt
539	1.7	292	2	H64129	probable protein-m
540	1.7	293	1	D69300	hemk protein (EC 1
					4-hydroxybenzoate
					succinyl-Coa synth

541	6	1.7	293	2	C86612	614	6	1.7	319	2	138138	protein-serine kin
542	6	1.7	293	2	B85585	615	6	1.7	319	2	C86036	probable transcrip
543	6	1.7	294	2	C69956	616	6	1.7	319	2	F85054	probable malonyl-C
544	6	1.7	294	2	A35266	617	6	1.7	319	2	A70638	hypothetical prote
545	6	1.7	294	2	T37999	618	6	1.7	319	2	T34525	hypothetical prote
546	6	1.7	295	2	H86907	619	6	1.7	320	2	S22615	hypothetical prote
547	6	1.7	295	2	T22049	620	6	1.7	321	2	B47043	hypothetical 35.9K
548	6	1.7	296	2	G86446	621	6	1.7	321	2	A56267	polyamine uptake s
549	6	1.7	296	2	F86839	622	6	1.7	321	2	A40603	epithelial cell at
550	6	1.7	297	2	F86836	623	6	1.7	321	2	H85990	probable dehydroge
551	6	1.7	297	2	S37636	624	6	1.7	322	2	T11101	iron(III) ABC tran
552	6	1.7	297	2	S22098	625	6	1.7	322	2	G72421	cation antiporter
553	6	1.7	297	2	A72130	626	6	1.7	322	2	E69009	hypothetical prote
554	6	1.7	297	2	E72102	627	6	1.7	322	2	T45862	oligopeptide trans
555	6	1.7	297	2	H86519	628	6	1.7	323	2	T45279	B-alpha pheromone
556	6	1.7	297	2	F85947	629	6	1.7	323	2	S61919	hypothetical prote
557	6	1.7	299	2	E69835	630	6	1.7	323	2	H70765	hypothetical prote
558	6	1.7	299	2	D81067	631	6	1.7	323	2	T32874	hypothetical prote
559	6	1.7	300	1	A32566	632	6	1.7	324	2	T04576	hypothetical prote
560	6	1.7	300	2	S47764	633	6	1.7	324	2	H84329	iron(III) ABC tran
561	6	1.7	300	2	B86028	634	6	1.7	326	1	C41335	microbial serine p
562	6	1.7	300	2	S15786	635	6	1.7	326	2	G64894	hypothetical prote
563	6	1.7	300	2	T06691	636	6	1.7	326	2	F85735	hypothetical prote
564	6	1.7	301	2	E81024	637	6	1.7	326	2	T50591	hypothetical prote
565	6	1.7	301	2	B82235	638	6	1.7	326	2	A86411	protein F3M18.6 li
566	6	1.7	301	2	F82235	639	6	1.7	326	2	E83773	probable peptidase
567	6	1.7	301	2	F86647	640	6	1.7	326	2	S36335	U2 snRNP 40K prote
568	6	1.7	302	2	F81016	641	6	1.7	326	2	C84885	cysteine synthase
569	6	1.7	302	2	H64026	642	6	1.7	327	2	F84290	DNA repair protein
570	6	1.7	302	2	S47526	643	6	1.7	327	2	S33672	chain length deter
571	6	1.7	303	2	D83082	644	6	1.7	327	2	F81317	UDP-glucose 4-epime
572	6	1.7	303	2	S28264	645	6	1.7	328	2	T30999	hypothetical prote
573	6	1.7	303	2	C72296	646	6	1.7	328	2	S44633	f22b7.8 protein -
574	6	1.7	303	2	H84216	647	6	1.7	328	2	D81861	conserved hypotet
575	6	1.7	303	2	B75428	648	6	1.7	328	2	T36963	hypothetical prote
576	6	1.7	304	2	F75632	649	6	1.7	329	2	H71461	delta-aminolevulin
577	6	1.7	305	2	F70533	650	6	1.7	329	2	T37891	hypothetical prote
578	6	1.7	305	2	H83518	651	6	1.7	329	2	T37695	probable potassium
579	6	1.7	305	2	D70432	652	6	1.7	330	2	T07394	conserved hypotet
580	6	1.7	305	2	D70432	653	6	1.7	331	2	E83368	transcription acti
581	6	1.7	306	2	D70250	654	6	1.7	331	2	T01824	probable Athlia re
582	6	1.7	307	2	G81905	655	6	1.7	333	2	A36925	hypothetical prote
583	6	1.7	307	2	B81111	656	6	1.7	333	2	E84476	hypothetical prote
584	6	1.7	307	2	G75280	657	6	1.7	333	2	T05984	hypothetical prote
585	6	1.7	308	2	JC2253	658	6	1.7	333	2	H65028	UDP-glucose 4-epim
586	6	1.7	308	2	C83978	659	6	1.7	334	2	D83788	hypothetical prote
587	6	1.7	308	2	T17248	660	6	1.7	334	2	E71103	hypothetical prote
588	6	1.7	309	1	CCB11H	661	6	1.7	334	2	T19637	probable thiosulfa
589	6	1.7	310	2	C39778	662	6	1.7	334	2	T04572	hypothetical prote
590	6	1.7	310	2	T29170	663	6	1.7	334	2	G85886	probable thiosulfa
591	6	1.7	311	2	T09656	664	6	1.7	335	2	H84365	hypothetical prote
592	6	1.7	311	2	B86211	665	6	1.7	335	2	H51106	peroxidase-like pr
593	6	1.7	311	2	T00087	666	6	1.7	335	2	T45730	hydroxymethylbilla
594	6	1.7	312	2	A64042	667	6	1.7	336	2	T39093	probable cysteine
595	6	1.7	312	2	T33160	668	6	1.7	336	2	T24387	homeobox protein c
596	6	1.7	313	2	T23536	669	6	1.7	337	2	T27635	conserved hypotet
597	6	1.7	313	2	T43044	670	6	1.7	337	2	F75599	probable nodulatio
598	6	1.7	313	2	F85205	671	6	1.7	338	2	E72636	oligopeptide trans
599	6	1.7	314	2	S58715	672	6	1.7	338	2	A53280	cathepsin L (EC 3.
600	6	1.7	314	2	JC5273	673	6	1.7	338	2	A86855	hypothetical prote
601	6	1.7	315	2	D70927	674	6	1.7	338	2	C75338	alcohol dehydrogen
602	6	1.7	315	2	A84634	675	6	1.7	339	2	A53810	flagellar switch p
603	6	1.7	315	2	C81962	676	6	1.7	339	2	T37487	myosin-related pro
604	6	1.7	316	1	MMW0V	677	6	1.7	339	2	T37030	hypothetical prote
605	6	1.7	317	2	S77827	678	6	1.7	340	2	B55544	hypothetical prote
606	6	1.7	317	2	T36926	679	6	1.7	340	2	T47733	hypothetical prote
607	6	1.7	317	2	T39736	680	6	1.7	341	2	S40612	hypothetical prote
608	6	1.7	318	2	S14948	681	6	1.7	341	2	G72667	hypothetical prote
609	6	1.7	318	2	B83527	682	6	1.7	342	2	T15541	hypothetical prote
610	6	1.7	318	2	S76700	683	6	1.7	342	2	S57165	transcriptional re
611	6	1.7	318	2	T35367	684	6	1.7	343	2	H83817	hypothetical prote
612	6	1.7	318	2	S25459	685	6	1.7	343	2	D69482	hypothetical prote
613	6	1.7	319	2	H65173	686	6	1.7	343	2		

687	6	1.7	344	2	T33057	hypothetical prote
688	6	1.7	344	2	T34835	probable transferr
689	6	1.7	345	2	T12361	NADH dehydrogenase
690	6	1.7	345	2	E69484	conserved hypothet
691	6	1.7	345	2	T25580	hypothetical prote
692	6	1.7	345	2	F70248	hypothetical prote
693	6	1.7	345	2	T49076	hypothetical prote
694	6	1.7	346	2	B75439	hypothetical prote
695	6	1.7	346	2	T48496	membrane protein -
696	6	1.7	347	2	T42528	hypothetical prote
697	6	1.7	347	2	T38998	probable galactosy
698	6	1.7	347	2	T38998	hypothetical prote
699	6	1.7	347	2	T38998	hypothetical prote
700	6	1.7	348	2	T51478	root cap 1 (RCP1)
701	6	1.7	348	2	T10946	peroxidase (EC 1.1
702	6	1.7	349	2	T17201	H+-transferring Ar
703	6	1.7	349	2	T15936	galactin receptor 1
704	6	1.7	349	2	T22346	hypothetical prote
705	6	1.7	350	2	S22456	hydroxyproline-ric
706	6	1.7	350	2	A05212	hypothetical prote
707	6	1.7	350	2	T43680	DNA repair protein
708	6	1.7	350	2	S51494	arabinogalactan en
709	6	1.7	351	2	S42546	photosystem II pro
710	6	1.7	352	2	T07229	photosystem II pro
711	6	1.7	352	2	A25466	photosystem II pro
712	6	1.7	352	2	S34495	photosystem II pro
713	6	1.7	352	2	T06935	photosystem II pro
714	6	1.7	352	2	E86430	photosystem II pro
715	6	1.7	352	2	C84364	H+-transferring Ar
716	6	1.7	353	1	F2NPD2	photosystem II pro
717	6	1.7	353	1	F2KRD2	photosystem II pro
718	6	1.7	353	1	F2LVD2	photosystem II pro
719	6	1.7	353	1	F2SPD2	photosystem II pro
720	6	1.7	353	1	F2PMD2	photosystem II pro
721	6	1.7	353	2	S00929	photosystem II pro
722	6	1.7	353	2	S03435	photosystem II pro
723	6	1.7	353	2	A31183	photosystem II pro
724	6	1.7	353	2	S58536	photosystem II pro
725	6	1.7	353	2	S52997	photosystem II pro
726	6	1.7	353	2	T07549	photosystem II pro
727	6	1.7	353	2	T27480	hypothetical prote
728	6	1.7	353	2	F84031	hypothetical prote
729	6	1.7	353	4	S52396	photosystem II pro
730	6	1.7	354	2	F72375	transcription anti
731	6	1.7	354	2	T47945	3-methyl-2-oxobuta
732	6	1.7	355	2	A75068	hypothetical prote
733	6	1.7	355	2	D84280	hypothetical prote
734	6	1.7	355	2	G75444	branched-chain ami
735	6	1.7	355	2	A44869	cell type-specific
736	6	1.7	357	2	T09966	farnesyltransferas
737	6	1.7	358	2	C84713	probable dioxygena
738	6	1.7	358	2	T38920	prostaglandin E2 r
739	6	1.7	358	2	S51312	Ep2 prostaglandin
740	6	1.7	358	2	H75084	hypothetical prote
741	6	1.7	358	2	H71058	hypothetical prote
742	6	1.7	358	2	T35104	probable two-compo
743	6	1.7	358	2	T29492	hypothetical prote
744	6	1.7	358	2	G70860	probable fecb - My
745	6	1.7	359	2	F83419	probable permease
746	6	1.7	360	2	D69623	flagellar biosynth
747	6	1.7	361	1	HLRB	MHC class I histoc
748	6	1.7	361	1	I46858	MHC class I RLA pr
749	6	1.7	362	2	S66674	prostaglandin E re
750	6	1.7	362	2	C81453	flagellar biosynth
751	6	1.7	362	2	H84494	hypothetical prote
752	6	1.7	362	2	T36079	hypothetical prote
753	6	1.7	364	2	T10945	peroxidase (EC 1.1
754	6	1.7	365	2	T22584	hypothetical prote
755	6	1.7	365	2	T06615	hypothetical prote
756	6	1.7	365	2	H69231	sensory transducti
757	6	1.7	366	2	T46249	hypothetical prote
758	6	1.7	366	2	S74387	hypothetical prote
759	6	1.7	366	2	T36035	ion transport prot



833	6	1.7	399	1	DXBHZ	protein z4 - barle	906	6	1.7	429	2	S49595	probable serine pr
834	6	1.7	399	2	T00631	hypothetical prote	907	6	1.7	429	2	D69335	H+/Na+-glutamate s
835	6	1.7	399	2	G75610	imidazolonepropi	908	6	1.7	430	2	A84887	probable phosphati
836	6	1.7	399	2	T30222	sensory protein ki	909	6	1.7	431	2	B82119	citrate synthase V
837	6	1.7	400	1	A46723	MAP kinase kinase	910	6	1.7	431	2	S76222	hypothetical prote
838	6	1.7	400	1	A48081	mitogen-activated	911	6	1.7	431	2	B75491	proton/sodium-glu
839	6	1.7	400	1	C82855	aromatic-amino-act	912	6	1.7	431	2	T36221	probable ATP-bindi
840	6	1.7	400	2	S02507	nitrogenase cofact	913	6	1.7	431	2	A84380	hypothetical prote
841	6	1.7	400	2	H72119	conserved hypothet	914	6	1.7	432	2	S08277	cyclin A - human
842	6	1.7	400	2	A86504	CT017 hypothetical	915	6	1.7	432	2	T16141	hypothetical prote
843	6	1.7	400	2	C70082	multidrug-efflux t	916	6	1.7	433	2	B71568	methy1-accepting c
844	6	1.7	401	1	152829	ptrrb protein - Esc	917	6	1.7	433	2	D82077	transport protein
845	6	1.7	401	2	S09626	hypothetical prote	918	6	1.7	435	1	S65982	probable integral
846	6	1.7	401	2	C81435	hypothetical prote	919	6	1.7	435	2	A81416	site-specific DNA-
847	6	1.7	401	2	D83022	probable emulsan r	920	6	1.7	437	2	JC1165	probable MFS trans
848	6	1.7	401	2	T44831	probable polygalac	921	6	1.7	437	2	H83175	glutamate-aspartat
849	6	1.7	402	2	C84529	probable polygalac	922	6	1.7	437	2	G86101	probable MFS trans
850	6	1.7	402	2	D84529	probable polygalac	923	6	1.7	437	2	A42384	glutamate-aspartat
851	6	1.7	402	2	T00995	probable polygalac	924	6	1.7	438	2	H83556	probable MFS trans
852	6	1.7	402	2	D82099	1-deoxy-D-xylulose	925	6	1.7	438	2	H82974	porin precursor, p
853	6	1.7	403	2	S02856	site-specific DNA-	926	6	1.7	438	2	S25260	hypothetical prote
854	6	1.7	403	2	T35370	probable transmemb	927	6	1.7	439	2	T22060	hipe protein - Pse
855	6	1.7	404	2	C70610	probable glucose-1	928	6	1.7	439	2	S61858	preprotein transio
856	6	1.7	404	2	H84827	probable polygalac	929	6	1.7	439	2	S27155	hypothetical prote
857	6	1.7	404	2	B84529	probable polygalac	930	6	1.7	439	2	T01807	hypothetical prote
858	6	1.7	404	2	A71548	hypothetical prote	931	6	1.7	439	2	B86168	hypothetical prote
859	6	1.7	404	2	T19858	sodium/dicarboxyla	932	6	1.7	440	2	H72784	probable alkaline
860	6	1.7	405	2	B82509	hypothetical prote	933	6	1.7	440	2	T40543	hypothetical prote
861	6	1.7	405	2	T19886	hypothetical prote	934	6	1.7	442	1	S11712	transcription init
862	6	1.7	406	2	S39865	hypothetical prote	935	6	1.7	442	1	T34661	hypothetical prote
863	6	1.7	406	2	D81964	diaminopimelate de	936	6	1.7	443	2	JC4978	oxidative stress p
864	6	1.7	406	2	T39786	hypothetical prote	937	6	1.7	443	2	I40793	dihydroliposamide S
865	6	1.7	407	2	S66260	metalloproteinase	938	6	1.7	443	2	T17320	hypothetical prote
866	6	1.7	408	2	T33939	probable alcohol d	939	6	1.7	443	2	D82847	membrane protein X
867	6	1.7	408	2	E70127	flagellar motor sw	940	6	1.7	444	1	C70349	DNA repair protein
868	6	1.7	409	2	A70563	probable transposa	941	6	1.7	444	2	T11474	NAOH dehydrogenase
869	6	1.7	410	1	C69585	aminopeptidase amp	942	6	1.7	444	2	A82962	proton-glutamate S
870	6	1.7	410	2	H72502	hypothetical prote	943	6	1.7	445	2	S77010	probable integral
871	6	1.7	410	2	D84356	probable acetyl-co	944	6	1.7	445	2	T37161	hypothetical prote
872	6	1.7	410	2	E70744	probable senX3 pro	945	6	1.7	445	2	S75859	probable two-compo
873	6	1.7	411	2	B84504	probable retroelem	946	6	1.7	445	2	A83245	hybrid cluster 14F
874	6	1.7	411	2	T49928	beta-glucosidase-1	947	6	1.7	446	2	E75075	hypothetical prote
875	6	1.7	412	1	S75358	coproporphyrinogen	948	6	1.7	446	2	C85343	probable aldehyde
876	6	1.7	413	2	A55338	transcription fact	949	6	1.7	447	2	H81848	aldehyde dehydroge
877	6	1.7	414	2	A81020	diaminopimelate de	950	6	1.7	447	2	D81093	hypothetical prote
878	6	1.7	415	2	T86279	probable transmemb	951	6	1.7	447	2	B96720	hypothetical prote
879	6	1.7	416	2	A55237	transcription fact	952	6	1.7	448	2	A83775	hypothetical prote
880	6	1.7	416	2	C39200	hypothetical prote	953	6	1.7	448	2	C96542	hypothetical prote
881	6	1.7	416	2	T23977	hypothetical prote	954	6	1.7	448	2	S57909	probable histidine
882	6	1.7	418	2	S72761	nitrogen fixation	955	6	1.7	448	2	T23263	hypothetical prote
883	6	1.7	419	1	S74500	zeaxanthin glucosy	956	6	1.7	449	2	C86095	aspartokinase III,
884	6	1.7	419	1	T45186	probable glucose-1	957	6	1.7	449	2	T47039	hypothetical prote
885	6	1.7	419	2	S76571	hypothetical prote	958	6	1.7	449	2	A47676	bacteriocin 28b -
886	6	1.7	419	2	F82183	probable outer mem	959	6	1.7	449	2	G70670	hypothetical prote
887	6	1.7	419	2	T33004	hypothetical prote	960	6	1.7	451	2	C84104	hypothetical prote
888	6	1.7	420	2	S684201	multidrug resistat	961	6	1.7	451	2	A96718	hypothetical prote
889	6	1.7	421	2	S26246	glutamate/aspartat	962	6	1.7	451	2	C83164	hypothetical prote
890	6	1.7	421	2	S26247	glutamate/aspartat	963	6	1.7	452	2	S74728	hypothetical prote
891	6	1.7	422	2	JC7080	melanin-concentrat	964	6	1.7	452	2	T00840	probable senescenc
892	6	1.7	422	2	E72042	conserved hypothet	965	6	1.7	452	2	A75484	conserved hypothet
893	6	1.7	423	1	F65681	CHLPN 76 kDa homol	966	6	1.7	453	2	S10840	gene ND4L, intron p
894	6	1.7	423	1	A29639	carboxypeptidase D	967	6	1.7	453	2	T39240	hypothetical prote
895	6	1.7	426	2	F70869	probable hemy' - M	968	6	1.7	454	1	T0U9BA	replication initia
896	6	1.7	426	2	T17336	hypothetical prote	969	6	1.7	454	2	T19433	hypothetical prote
897	6	1.7	427	1	VWEC	histidinol dehydro	970	6	1.7	456	2	F84724	conserved glucosyl
898	6	1.7	427	2	A65641	citrate synthase I	971	6	1.7	456	2	D82810	conserved glucosyl
899	6	1.7	427	2	H85572	hypothetical prote	972	6	1.7	456	2	B86834	sensor protein kin
900	6	1.7	427	2	H72678	hypothetical prote	973	6	1.7	456	2	T48291	preprotein translo
901	6	1.7	428	2	S22377	aspartate transami	974	6	1.7	457	2	E81663	hypothetical prote
902	6	1.7	428	2	S22379	hypothetical prote	975	6	1.7	457	2	T13451	hypothetical prote
903	6	1.7	428	2	S27898	hypothetical prote	976	6	1.7	457	2	T40770	colled coli protei
904	6	1.7	428	2	E84192	zinc metalloprotei	977	6	1.7	458	2	JC2467	biphenyl dioxygena
905	6	1.7	428	2	A83807	hypothetical prote	978	6	1.7	458	2	A39431	phenylalanine tran

```

979 6 1.7 458 2 A85556 phenylalanine-spec
980 6 1.7 459 2 F84935 argininosuccinate
981 6 1.7 459 2 D69852 sodium/galactoside
982 6 1.7 459 2 C81750 conserved hypotet
983 6 1.7 459 2 B86176 protein Flp19.10
984 6 1.7 459 2 B72361 conserved hypotet
985 6 1.7 459 2 D70914 conserved hypotet
986 6 1.7 460 2 S47964 ginseng - ginkgo
987 6 1.7 461 2 A81885 NAD(P) transhydrog
988 6 1.7 461 2 C81137 lipid A biosynthes
989 6 1.7 462 2 E81551 hypothetical prote
990 6 1.7 463 2 A64932 part of a kinase I
991 6 1.7 463 2 C85781 cysteine proteins
992 6 1.7 464 2 S24802 probable hydrolyti
993 6 1.7 464 2 T35943 conjugal transfer
994 6 1.7 465 2 E82865 myosin alpha heavy
995 6 1.7 465 2 A02986 probable membrane
996 6 1.7 465 2 S34929 probable amino aci
997 6 1.7 465 2 A83046 adenylosuccinate I
998 6 1.7 467 2 T04540 acyltransferase -
999 6 1.7 467 2 B72119 probable NADH dehy
1000 6 1.7 468 2 A72619

```

## ALIGNMENTS

```

RESULT 1
T08667
hypothetical protein DKFZp54700510.1 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 23-Jul-1999
C:Accession: T08667
R:Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, May 1999
A:Reference number: Z16466
A:Accession: T08667
A:Molecule type: mRNA
A:Residues: 1-330 <R0D>
A:Cross-references: EMBL:AL049929
A:Experimental source: fetal brain; clone DKFZp54700510
C:Genetics:
A:Note: DKFZp54700510.1

```

```

Query Match          75.4%: Score 264; DB 2; Length 330;
Best Local Similarity 100.0%: Pred. No. 4.4e-254;
Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 21 ILKPSGVFRNGNMPFGERIPDVAAISMGFSYKEDLSMPGLAVGNLFHRRPRTVVMV 80
    |||||||
DB 1 ILKPSGVFRNGNMPFGERIPDVAAISMGFSYKEDLSMPGLAVGNLFHRRPRTVVMV 60
QY 81 KGVNKLALPPGVSIVYLENNAVPPSLDSVAANSIHSLFSEETPVVQLAPSEERYMCKA 140
    |||||||
DB 61 KGVNKLALPPGVSIVYLENNAVPPSLDSVAANSIHSLFSEETPVVQLAPSEERYMCKA 120
QY 141 NSVEEDLSYTLQNLNRLFDQNSVLSPLNSLSNNNEVDLFLSELQVLDHISLSLRH 200
    |||||||
DB 121 NSVEEDLSYTLQNLNRLFDQNSVLSPLNSLSNNNEVDLFLSELQVLDHISLSLRH 180
QY 201 KHLAKHSDPLSLSLAGLDIGKRGYGESEQFRASKIIVDLAKFADMSLYGNAV 260
    |||||||
DB 181 KHLAKHSDPLSLSLAGLDIGKRGYGESEQFRASKIIVDLAKFADMSLYGNAV 240
QY 261 VELVTVKSDYSLIKRTTILEAK 284
    |||||||
DB 241 VELVTVKSDYSLIKRTTILEAK 264

```

```

RESULT 2
T44797
iron transport membrane protein ItrpB [Imported] - Corynebacterium diphtheriae

```

```

C:Species: Corynebacterium diphtheriae
C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 18-Feb-2000
C:Accession: T44797
R:Qian, Y.; Holmes, R.K.
submitted to the EMBL Data Library, August 1999
A:Description: ABC transporters involved in iron uptake in Corynebacterium diphtheriae
A:Reference number: Z22846
A:Accession: T44797
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-343 <Q1A>
A:Cross-references: EMBL:AF176902; P1DN:AMD53126.1
A:Experimental source: strain C7
C:Genetics:
A:Gene: ItrpB
C:Superfamily: vitamin B12 transport protein btuc

```

```

Query Match          2.6%: Score 9; DB 2; Length 343;
Best Local Similarity 100.0%: Pred. No. 0.95;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 VFVVLALV 11
    |||||||
DB 21 VFVVLALV 29

```

```

RESULT 3
T36132
hypothetical protein SCE19A.09 - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
R:Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1999
A:Reference number: Z21598
A:Accession: T36132
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-55 <SEE>
A:Cross-references: EMBL:AL096852; P1DN:CAB50990.1; GSPDB:GN00070; SCOEDB:SCE19A.09
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCE19A.09

```

```

Query Match          2.3%: Score 8; DB 2; Length 55;
Best Local Similarity 100.0%: Pred. No. 1.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 VVLLALVA 12
    |||||||
DB 12 VVLLALVA 19

```

```

RESULT 4
B84227
hypothetical protein Vng0692c [Imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: B84227
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky,
J.; Lettauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.;
A>Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; M01D:20504483
A:Accession: B84227
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-156 <STO>
A:Cross-references: GB:AE004437; NID:g10580278; P1DN:AMG19182.1; GSPDB:GN00138

```

C:Genetics:  
A:Gene: VMS0692C

Query Match 2.3%; Score 8; DB 2; Length 156;  
Best Local Similarity 100.0%; Pred. No. 4.7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 LIALVAGV 14  
|||||  
DB 63 LIALVAGV 70

RESULT 5  
F86728  
beta-subunit of L-serine dehydratase [imported] - Lactococcus lactis subsp. lactis (str)

C:Species: Lactococcus lactis subsp. lactis  
C>Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 23-Mar-2001  
C:Accession: F86728  
R:Boletín, A.; Wincker, P.; Mauger, S.; Jallion, O.; Malarne, K.; Weissenbach, J.; Ehrlich  
Genome Res. In press, 2001  
A:Title: The complete genome sequence of the lactic acid bacterium.  
A:Reference number: A86625  
A:Accession: F86728

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-223 <STO>  
A:Cross-references: GB:AE005176; MID:g12723754; PIDN:AAK04928.1; GSPDB:GN00146  
A:Experimental source: strain IL1403  
C:Genetics:  
A:Gene: sdab

Query Match 2.3%; Score 8; DB 2; Length 223;  
Best Local Similarity 100.0%; Pred. No. 6.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 ALVAGVIG 16  
|||||  
DB 66 ALVAGVIG 73

RESULT 6  
T23876  
hypothetical protein R03E1.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T23876  
R:McMurray, A.  
submitted to the EMBL Data Library, March 1997  
A:Reference number: Z19812  
A:Accession: T23876

A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-324 <WIL>  
A:Cross-references: EMBL:Z92837; PIDN:CA07401.1; GSPDB:GN00028; CESP:R03E1.2  
A:Experimental source: clone R03E1  
C:Genetics:  
A:Gene: CESP:R03E1.2  
A:Map position: X  
A:Introns: 59/1; 180/2; 293/3

Query Match 2.3%; Score 8; DB 2; Length 324;  
Best Local Similarity 100.0%; Pred. No. 8.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 336 DSIIYRMT 343  
|||||  
DB 310 DSIIYRMT 317

RESULT 7

F86256  
hypothetical protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: F86256  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso  
Chili, C.W.; Chung, M.K.; Comp. L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Hultzer, L.

Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzla  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: F86256

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-468 <STO>  
A:Cross-references: GB:AE005172; MID:g10086517; PIDN:AAG12577.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 2.3%; Score 8; DB 2; Length 468;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 GSVFRNG 33  
|||||  
DB 261 GSVFRNG 268

RESULT 8

T10568  
probable serine/threonine-specific protein kinase (EC 2.7.1.-) F25E4.100 - Arabidopsi  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jan-2000  
C:Accession: T10568  
R:Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Ban  
submitted to the Protein Sequence Database, June 1999

A:Reference number: Z16533  
A:Accession: T10568  
A:Molecule type: DNA  
A:Residues: 1-656 <BEV>  
A:Cross-references: EMBL:AL050399; GSPDB:GN00062; ATSP:F25E4.100  
A:Experimental source: cultivar Columbia; BAC clone F25E4  
C:Genetics:  
A:Gene: ATSP:F25E4.100

A:Map position: 4  
A:Introns: 297/1; 337/3; 416/1; 495/2; 546/3  
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom  
C:Keywords: phosphotransferase; protein kinase  
F:319-608/Domain: protein kinase homology KIN>

Query Match 2.3%; Score 8; DB 2; Length 656;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 FVVLALV 11  
|||||  
DB 273 FVVLALV 280

RESULT 9

DB4481  
probable retroelement pol polyprotein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)  
C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: DB4481

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y

M.: Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.; euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.

A:Reference number: A84420; MUID:20083487

A:Accession: D84481

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1413 <STO>

A:Cross-references: GB:AE002093; NID:93779026; PIDN:AA67205.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g07010

A:Map position: 2

C:Superfamily: retrovirus-related polyprotein

Query Match 2.3%; Score 8; DB 2; Length 1413;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 163 SVLSLPL 170  
|||||||  
Db 523 SVLSLPL 530

RESULT 10

C84512

Probable retroelement pol polyprotein [imported] - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cress)

C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001

C:Accession: C84512

R:Lin. X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; eus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.

A:Reference number: A84420; MUID:20083487

A:Accession: C84512

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1501 <STO>

A:Cross-references: GB:AE002093; NID:94388829; PIDN:AA19784.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g13940

A:Map position: 2

C:Superfamily: retrovirus-related polyprotein

Query Match 2.3%; Score 8; DB 2; Length 1501;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 163 SVLSLPL 170  
|||||||  
Db 540 SVLSLPL 547

RESULT 11

C40228

neurexin II-alpha precursor - rat

C:Species: *Rattus norvegicus* (Norway rat)

C>Date: 21-Jul-1995 #sequence\_revision 21-Jul-1995 #text\_change 11-Jan-2000

C:Accession: C40228; S27886; S27887

R:Ushakov, Y.A.; Petrenko, A.G.; Geppert, M.; Suedhof, T.C. Science 257, 50-56, 1992

A:Title: Neurexins: synaptic cell surface proteins related to the alpha-latrotoxin receptor

A:Reference number: A40228; MUID:92320296

A:Accession: C40228

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1715 <OSH>

A:Cross-references: GB:M96376; NID:9205714; PIDN:AAA1707.1; PID:9205716

A:Note: authors translated the codon GAC for residue 1464 as Thr and ACC for residue 146

R:Ushakov, Y.A.; Petrenko, A.G.; Geppert, M.; Sudhof, T.C. Submitted to the EMBL Data Library, July 1992

A:Description: Neurexins: Synaptic cell surface proteins related to the alpha-latrotoxin

A:Reference number: S27884

A:Accession: S27886

A:Molecule type: mRNA

A:Residues: 1-1666, 'CRK', 1670, 'PREKLLPG', 1683-1685, 'GL', 1688, 'LDLA', 1694-1695, 'CCVCR

C:Cross-references: EMBL:M96376; NID:9205714; PIDN:AAA1706.1; PID:9205715

C:Genetics:

A:Introns: 1666/2

C:Superfamily: unassigned EGF-related proteins; EGF homology

C:Keywords: alternative splicing; transmembrane protein

F:1-27/Domain: signal sequence #status predicted <STR>

F:28-1715/Product: neurexin II-alpha #status predicted <MAT>

F:694-726/Domain: EGF homology <EGF>

F:1103-1135/Domain: EGF homology <EGF1>

Query Match 2.3%; Score 8; DB 2; Length 1715;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 LLALVAGY 14  
|||||||  
Db 19 LLALVAGY 26

RESULT 12

F83345

Probable non-ribosomal peptide synthetase PA2402 [imported] - *Pseudomonas aeruginosa*

C:Species: *Pseudomonas aeruginosa*

C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: F83345

R:Stover, C.K.; Pham, X.Q.; Ervin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lartig, K.; L. Nature 406, 959-964, 2000

A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen

A:Reference number: A82950; MUID:20437337

A:Accession: F83345

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-5149 <SMO>

A:Cross-references: GB:AE004667; GB:AE004091; NID:99948444; PIDN:AA605790.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA2402

C:Superfamily: acyl carrier protein homology; acetate-CoA ligase homology

C:Keywords: carrier protein; phosphopantetheine; phosphoprotein

F:557-998/Domain: acetate-CoA ligase homology <ACT1>

F:1015-1082/Domain: acyl carrier protein homology <ACP1>

F:2034-2479/Domain: acetate-CoA ligase homology <ACP1>

F:2496-2564/Domain: acyl carrier protein homology <ACP2>

F:3084-3525/Domain: acetate-CoA ligase homology <ACP2>

F:3581-3608/Domain: acyl carrier protein homology <ACP3>

F:4583-5040/Domain: acetate-CoA ligase homology <ACP3>

F:5057-5125/Domain: acyl carrier protein homology <ACP4>

F:1047, 2528, 3573, 5089/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 2.3%; Score 8; DB 2; Length 5149;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 164 VLSSLPLN 171  
|||||||  
Db 2462 VLSSLPLN 2469

RESULT 13

C54037

Splicing regulatory protein SWAP homolog (alternatively spliced, clone pFL2) - human

C:Species: *Homo sapiens* (man)

C>Date: 06-Oct-1994 #sequence\_revision 18-Nov-1994 #text\_change 28-Feb-1997  
 C:Accession: C54037  
 R:Denhez, F.; Lafyatis, R.  
 J. Biol. Chem. 269, 16170-16179, 1994  
 A:Title: Conservation of regulated alternative splicing and identification of functional  
 A:Reference number: A54037; MUID:9426805  
 A:Accession: C54037  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid  
 A:Residues: 1-29 <DEN>  
 A:Experimental source: liver  
 A:Note: sequence extracted from NCBI backbone (NCBIF:149928)

Query Match 2.0%; Score 7; DB 2; Length 29;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 281 LEAKOK 287  
 |||||  
 DB 8 LEAKOK 14

RESULT 14  
 S42862  
 hypothetical protein 73 - rice  
 C:Species: Oryza sativa (rice)  
 C>Date: 07-Sep-1994 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999  
 C:Accession: S42862  
 R:Akagi, H.; Sakamoto, M.; Shinjo, C.; Shimada, H.; Fujimura, T.  
 Curr. Genet. 29, 52-58, 1994  
 A:Title: A unique sequence located downstream from the rice mitochondrial atp6 may cause  
 A:Reference number: S42858; MUID:94363737  
 A:Accession: S42862  
 A:Status: preliminary; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-73 <AKA>  
 A:Cross-references: GB:D14339; NID:g440856; PID:d1003782; PID:g440861

Query Match 2.0%; Score 7; DB 2; Length 73;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 102 VPSIDS 108  
 |||||  
 DB 50 VPSIDS 56

RESULT 15  
 T2665  
 hypothetical protein Y38E10A.o - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T2665  
 R:Wallis, J.  
 submitted to the EMBL Data Library, September 1999  
 A:Reference number: Z20252  
 A:Accession: T2665  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-74 <WII>  
 A:Cross-references: EMBL:AL110484; NID:e1542205; PIDN:CA54406.1; CESP:Y38E10A.o  
 A:Experimental source: clone Y38E10A  
 A:Genetics:  
 A:Gene: CESP:Y38E10A.o  
 A:Introns: 35/1

Query Match 2.0%; Score 7; DB 2; Length 74;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 VFVVLLA 9  
 |||||  
 DB 10 VFVVLLA 16

RESULT 16  
 T44129  
 hypothetical protein [imported] - Staphylococcus aureus (fragment)  
 C:Species: Staphylococcus aureus  
 C>Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 11-May-2000  
 C:Accession: T44129  
 R:Ito, T.; Katayama, Y.; Hiramatsu, K.  
 Antimicrob. Agents Chemother. 43, 1449-1458, 1999  
 A:Title: Cloning and nucleotide sequence determination of the entire mec DNA of pre-m  
 A:Reference number: Z22733; MUID:99278010  
 A:Accession: T44129  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-109 <ITO>  
 A:Cross-references: EMBL:D86934; PIDN:BAAB2231.1  
 A:Experimental source: strain N315

Query Match 2.0%; Score 7; DB 2; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 163 SVLSLSP 169  
 |||||  
 DB 59 SVLSLSP 65

RESULT 17  
 A45069  
 omega-agatoxin IA precursor - funnel-weaving spider (Agelenopsis aperta)  
 N:Alternate names: Ca2+ channel blocking toxin  
 C:Species: Agelenopsis aperta  
 C>Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 26-Jan-1996  
 C:Accession: A45069; A34923  
 R:Sanjos, A.D.; Imperial, J.S.; Chaudhary, T.; Beavis, R.C.; Chalf, B.T.; Hunsperger, J.  
 J. Biol. Chem. 267, 20701-20705, 1992  
 A:Title: Heterodimeric structure of the spider toxin omega-agatoxin IA revealed by pr  
 A:Reference number: A45069; MUID:93015969  
 A:Accession: A45069  
 A:Molecule type: mRNA; protein  
 A:Residues: 1-112 <SAN>  
 A:Experimental source: venom duct  
 A:Note: sequence extracted from NCBI backbone (NCBIN:116031, NCBIF:116032)  
 A:Note: the carboxyl-terminal tripeptide is disulfide linked with the remaining porti  
 R:Adams, M.E.; Bindokas, V.P.; Hasegawa, L.; Venema, V.J.  
 J. Biol. Chem. 265, 861-867, 1990  
 A:Title: omega-Agatoxins: novel calcium channel antagonists of two subtypes from funn  
 A:Reference number: A34923; MUID:90110147  
 A:Accession: A34923  
 A:Molecule type: protein  
 A:Residues: 37-100, 'N', 102 <ADA>  
 A:Experimental source: signal sequence  
 A:Keywords: acetylcholine release inhibition; calcium channel inhibitor; presynaptic  
 F:1-19/Domain: signal sequence #status predicted <STG>  
 F:37-102, 110-112/Product: omega-agatoxin IA #status experimental <MAT>

Query Match 2.0%; Score 7; DB 2; Length 112;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 87 ALPPGSV 93  
 |||||  
 DB 39 ALPPGSV 45

RESULT 18  
 G65203  
 y1jd protein - Escherichia coli (strain K-12)

C:Species: Escherichia coli  
 C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 08-Oct-1999  
 C:Accession: G65203  
 R:Battner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co-  
 .A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:97426617  
 A:Accession: G65203  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-119 <BLAT>  
 A:Cross-references: GB:AE000470; GB:U00096; NID:q2367332; PIDN:AACT6946.1; PID:q1790402;  
 A:Experimental source: strain K-12, substrain MG1655  
 C:Genetics:  
 A:Gene: y1jd

Query Match 2.0%; Score 7; DB 2; Length 119;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LIALVAG 13  
 |||||  
 Db 12 LIALVAG 18

## RESULT 19

D86088  
 hypothetical protein y1jd [Imported] - Escherichia coli (strain 0157:H7)  
 C:Species: Escherichia coli  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001

C:Accession: D86088  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Petramousis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli 0157:H7.  
 A:Reference number: A85480; MUID:21074935; PMID:11206551  
 A:Accession: D86088  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-119 <STO>  
 A:Cross-references: GB:AE005174; NID:q12518891; PIDN:AA659168.1; GSPDB:GN00145; UMGCP:255  
 A:Experimental source: strain 0157:H7, substrain EDL933  
 C:Genetics:  
 A:Gene: y1jd

Query Match 2.0%; Score 7; DB 2; Length 119;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LIALVAG 13  
 |||||  
 Db 12 LIALVAG 18

## RESULT 20

S21565  
 hypothetical protein B - Escherichia coli

C:Species: Escherichia coli  
 C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 08-Oct-1999  
 C:Accession: S21565  
 R:Gustafsson, C.  
 submitted to the EMBL Data Library, May 1992  
 A:Reference number: S21565  
 A:Accession: S21565  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-120 <GUS>  
 A:Cross-references: EMBL:X66026; NID:q43236; PIDN:CAA46824.1; PID:q43239

Query Match 2.0%; Score 7; DB 2; Length 120;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LIALVAG 13  
 |||||  
 Db 13 LIALVAG 19

## RESULT 21

B64208  
 hypothetical protein MG074 - Mycoplasma genitalium

C:Species: Mycoplasma genitalium  
 C:Date: 10-Nov-1995 #sequence\_revision 10-Nov-1995 #text\_change 21-Jul-2000  
 C:Accession: B64208  
 R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.  
 M.; Fuhrmann, J.; Nguyen, D.; Uitterback, T.R.; Saudek, D.W.; Phillips, C.A.; Merrick,  
 C.A.; Venter, J.C.  
 Science 270, 397-403, 1995  
 A:Title: The minimal gene complement of Mycoplasma genitalium.  
 A:Reference number: A64200; MUID:96026346  
 A:Accession: B64208  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-137 <RIG>  
 A:Cross-references: GB:U039687; GB:L43967; NID:q3844663; PIDN:AACT1292.1; PID:q1045750  
 A:Experimental source: strain G-37  
 C:Genetics:  
 A:Genetic code: SGC3  
 C:Superfamily: hypothetical protein MG074

Query Match 2.0%; Score 7; DB 2; Length 137;  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 LILFSEL 187  
 |||||  
 Db 14 LILFSEL 20

## RESULT 22

S14025  
 hypothetical protein 31 - Chlamydomonas reinhardtii transposon

C:Species: Chlamydomonas reinhardtii  
 C:Date: 21-Nov-1993 #sequence\_revision 01-Dec-1995 #text\_change 21-Jul-2000  
 C:Accession: S14025  
 R:Day, A.; Rochaix, J.D.  
 Nucleic Acids Res. 19, 1259-1266, 1991  
 A:Title: A transposon with an unusual ITR arrangement from Chlamydomonas reinhardtii  
 A:Reference number: S14018; MUID:91232906  
 A:Accession: S14025  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-137 <DAY>  
 A:Cross-references: EMBL:X56231; NID:q18229; PIDN:CAA39687.1; PID:q1360722  
 C:Genetics:  
 A:Mobile element: transposon

Query Match 2.0%; Score 7; DB 2; Length 137;  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 VLQAPS 130  
 |||||  
 Db 17 VLQAPS 23

## RESULT 23

D69549  
 Iron-dependent repressor homolog - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
 C:Accession: D69549  
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
 R: Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirschner, E.F.;  
 Glodok, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
 Nature 390, 364-370, 1997  
 A:Authors: Overbeek, R.; Cotton, M.D.; Spriggs, T.; Atliach, P.; Kalne, B.P.; Sykes, S.  
 Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A>Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae  
 A:Reference number: A69250; MUID:98049343  
 A:Accession: D69549  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-148 <KLE>  
 A:Cross-references: GB:AE001111; GB:AE000782; NID:g2689434; PIDN:AAB91276.1; PID:g265070  
 C:Superfamily: conserved hypothetical protein MJ0568

Query Match 2.0%; Score 7; DB 1; Length 148;  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 TLRLRLN 156  
 DB 109 TLRLRLN 115

## RESULT 24

C83091  
 hypothetical protein PA4441 [Imported] - Pseudomonas aeruginosa (strain PA01)  
 C:Species: Pseudomonas aeruginosa  
 C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: C83091  
 R:Stover, C.K.; Pham, X.Q.; Errin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B  
 adman, S.; Yuan, T.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,  
 .; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 A:Reference number: A82950; MUID:20437337  
 A:Accession: C83091  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-148 <STO>  
 A:Cross-references: GB:AE004858; GB:AE004091; NID:g9950668; PIDN:AAG07829.1; GSPDB:GN001  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA4441

Query Match 2.0%; Score 7; DB 2; Length 148;  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LALVAGV 14  
 DB 13 LALVAGV 19

## RESULT 25

G69014  
 conserved hypothetical protein MTH110 - Methanobacterium thermoautotrophicum (strain DE  
 C:Species: Methanobacterium thermoautotrophicum  
 C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 24-Sep-1999  
 C:Accession: G69014  
 R:Smith, D.R.; Donette-Stamm, L.A.; Delouhery, C.; Lee, H.; Dubois, J.; Aldredge, T.;  
 Qiu, D.; Spadefora, R.; Viscite, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.  
 K.; S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
 J. Bacteriol. 179, 7135-7155, 1997  
 A>Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct  
 A:Reference number: A69000; MUID:98037514  
 A:Accession: G69014  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA

A:Residues: 1-151 <MTH>  
 A:Cross-references: GB:AE000881; GB:AE000666; NID:g2622206; PIDN:AAB85599.1; PID:g262  
 A:Experimental source: strain Delta H  
 C:Genetics:  
 A:Gene: MTH110  
 A:Start codon: GTG  
 C:Superfamily: conserved hypothetical protein MJ1244

Query Match 2.0%; Score 7; DB 2; Length 151;  
 Best Local Similarity 100.0%; Pred. No. 45;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 GFSVKED 57  
 DB 87 GFSVKED 93

## RESULT 26

C:Phycocyanin alpha chain - Synechococcus sp. (PCC 6301)  
 C:Species: Synechococcus sp.  
 A:Variety: PCC 6301  
 A:Note: Synechococcus sp. PCC 6301 (ATCC 27144) was named Anacystis nidulans in other  
 C>Date: 18-Dec-1981 #sequence\_revision 02-Jul-1996 #text\_change 11-Jun-1999  
 C:Accession: A29015; S60070; S60075; A00317  
 R:Lau, R.H.; Alvarado-Urbina, G.; Lau, P.C.K.  
 Gene 52, 21-29, 1987  
 A>Title: Phycocyanin alpha-subunit gene of Anacystis nidulans R2: cloning, nucleotide  
 A:Reference number: A29015; MUID:87248092  
 A:Note: Anacystis nidulans  
 A:Accession: A29015  
 A:Molecule type: DNA  
 A:Residues: 1-163 <LAU>  
 A:Cross-references: GB:M16325; NID:g142137; PIDN:AAA22051.1; PID:g142139  
 A:Experimental source: strain R2  
 R:Kalla, S.R.; Lind, L.K.; Lidholm, J.; Gustafsson, P.  
 J. Bacteriol. 170, 2961-2970, 1988  
 A>Title: Transcriptional organization of the phycocyanin subunit gene clusters of the  
 A:Reference number: S60069; MUID:88257006  
 A:Accession: S60070  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-163 <KAL>  
 A:Cross-references: EMBL:M94218; NID:g142121; PIDN:AAA64527.1; PID:g142123  
 A:Experimental source: PCC 6301  
 A:Note: source designated as Anacystis nidulans  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1992  
 A:Accession: S60075  
 A>Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-163 <KA2>  
 A:Cross-references: EMBL:M94218; NID:g142121; PIDN:AAA64527.1; PID:g142123  
 A:Note: the source is designated as Anacystis nidulans  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1992  
 R:Walsh, R.G.; Wingfield, P.; Glazer, A.N.; Delange, R.J.  
 Fed. Proc. 39, 1998, 1980  
 A:Reference number: A00317  
 A:Accession: A00317  
 A:Molecule type: protein  
 A:Residues: 2-35, D, 55-120, TK, 123-157, IL, 160-163 <MAL>  
 C:Complex: associates with beta chain (see PIR:CFCYB)  
 C:Function:  
 A:Description: photon energy transfer from phycoerythrin to allophycocyanin  
 A:Pathway: photosynthesis  
 C:Superfamily: phycocyanin  
 C:Keywords: chromoprotein; photosynthesis; phycocyanobilin  
 P:2-163/Product: C-phycocyanin alpha chain #status experimental <MAT>  
 P:85/Binding site: phycocyanobilin (Cys) (covalent) #status experimental

Query Match 2.0%; Score 7; DB 1; Length 163;  
 Best Local Similarity 100.0%; Pred. No. 48;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 216 LAGLDEI 222  
 Db 113 LAGLDEI 119

# RESULT 27

675525

conserved hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000

C:Accession: G75525

R:White, O.; Eelsen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896

A:Accession: G75525

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-168 <NH1>

A:Cross-References: GB:AE001898; GB:AE00513; NID:96458057; PIDN:AAE09961.1; PID:9645804

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR0380

A:Map position: 1

Query Match 2.0%; Score 7; DB 2; Length 168;  
 Best Local Similarity 100.0%; Pred. No. 50;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 LALAV 321  
 Db 58 LALAV 64

RESULT 28

69467

conserved hypothetical protein AFI740 - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 29-Sep-1999

C:Accession: C69467

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson

; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirschner, E.F.

Nature 390, 364-370, 1997

A:Authors: Uetzel, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes, S.

Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae

A:Reference number: A69250; MUID:98049343

A:Accession: C69467

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-179 <KLE>

A:Cross-References: GB:AE000982; GB:AE000782; NID:92689305; PIDN:AA89505.1; PID:9264880

C:Superfamily: conserved hypothetical protein MJ1600

Query Match 2.0%; Score 7; DB 2; Length 179;  
 Best Local Similarity 100.0%; Pred. No. 53;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 318 LALAVIT 324  
 Db 152 LALAVIT 158

RESULT 29

C72495

hypothetical protein APE2602 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix

C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jun-2000

C:Accession: C72495

R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta

awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.

DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero

A:Reference number: A72450; MUID:99310339

A:Accession: C72495

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-184 <KAM>

A:Cross-References: DDBJ:AP00064; NID:95105945; PIDN:BA81619.1; PID:01045405; PID:9

A:Experimental source: strain K1

C:Genetics:

A:Gene: APE2602

C:Superfamily: Aeropyrum pernix hypothetical protein APE2602

Query Match 2.0%; Score 7; DB 2; Length 184;  
 Best Local Similarity 100.0%; Pred. No. 54;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 317 LALAVIT 323  
 Db 15 LALAVIT 21

RESULT 30

S28739

hypothetical protein 2 - Halobacterium salinarum phage phi-H insertion sequence ISH1.

C:Species: Halobacterium salinarum phage phi-H

C>Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 21-Jul-2000

C:Accession: S28739

R:Schubert, H.; Palm, P.; Dick, K.; Gramp, B.

EMBO J. 3, 1717-1722, 1984

A:Title: Sequence analysis of the insertion element ISH1.8 and of associated structur

A:Reference number: S28732

A:Accession: S28739

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-195 <SCH>

A:Cross-References: EMBL:X00805; NID:915525; PIDN:CAA25389.1; PID:9579187

A:Note: the source is designated as phage phi-H of the archaeobacterium Halobacterium

C:Genetics:

A:Mobile element: insertion sequence ISH1.8

A:Start codon: GTG

C:Superfamily: Halobacterium halobium phage phi-H insertion sequence ISH1.8 hypotacti

Query Match 2.0%; Score 7; DB 2; Length 195;  
 Best Local Similarity 100.0%; Pred. No. 57;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 IPDVAAL 48  
 Db 33 IPDVAAL 39

RESULT 31

T29016

hypothetical protein ZK84.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T29016

R:Kirsten, J.

submitted to the EMBL Data Library, April 1995

A:Description: The sequence of C. elegans cosmid ZK84.

A:Reference number: Z20553

A:Accession: T29016

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-196 <KIR>



A:Cross-references: EMBL:U23181; PIDN:AAC48205.1; GSPDB:GN00020; CESP:ZK84.2  
 A:Experimental source: strain Bristol N2; clone ZK84

C:Genetics:  
 A:Gene: CESP:ZK84.2  
 A:Map position: 2  
 A:introns: 50/1; 147/1

Query Match 2.0%; Score 7; DB 2; Length 196;  
 Best Local Similarity 100.0%; Pred. No. 57;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 84 NKIALP 90  
 |||||||  
 Db 162 NKIALP 168

RESULT 32  
 S57836  
 Lymphoid-specific interferon regulator factor - mouse (fragments)

C:Species: Mus musculus (house mouse)  
 C:Date: 28-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 17-Mar-1999  
 C:Accession: S57836  
 R:Matuyama, T.; Grossman, A.; Mitternacker, H.W.; Siderovski, D.P.; Kiefer, F.; Kawakami  
 Nucleic Acids Res. 23, 2127-2136, 1995  
 A:Title: Molecular cloning of LSIF, a lymphoid-specific member of the interferon regula  
 A:Reference number: S57836; MUID:95334364  
 A:Accession: S57836  
 A:Status: preliminary; nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-203 <MAT>  
 A:Cross-references: EMBL:U11692  
 C:Superfamily: lymphoid-specific interferon regulatory factor

Query Match 2.0%; Score 7; DB 2; Length 203;  
 Best Local Similarity 100.0%; Pred. No. 59;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 183 FLSELOV 189  
 |||||||  
 Db 174 FLSELOV 180

RESULT 33

T45907  
 hypothetical protein F4P12.370 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
 C:Accession: T45907  
 R:Bioecker, H.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quetier, F.; Salanoubat, M.  
 submitted to the Protein Sequence Database, January 2000  
 A:Reference number: Z23016  
 A:Accession: T45907  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-203 <BLO>  
 A:Cross-references: EMBL:AL132966  
 A:Experimental source: cultivar Columbia; BAC clone F4P12

C:Genetics:  
 A:Map position: 3  
 A:introns: 96/3  
 A:Note: F4P12.370

Query Match 2.0%; Score 7; DB 2; Length 203;  
 Best Local Similarity 100.0%; Pred. No. 59;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 167 SLPLNSL 173  
 |||||||  
 Db 70 SLPLNSL 76

RESULT 34  
 F70476  
 conserved hypothetical protein aq\_2059 - Aquifex aeolicus

C:Species: Aquifex aeolicus  
 C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 17-Mar-2000  
 C:Accession: F70476  
 R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;  
 V.  
 Nature 392, 353-358, 1998  
 A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
 A:Reference number: A70300; MUID:98196666  
 A:Accession: F70476  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-213 <AOP>  
 A:Cross-references: GB:AE000770; NID:g2984274; PIDN:AAC07809.1; PID:g2984284; GB:AE00  
 A:Experimental source: strain VFS  
 C:Genetics:  
 A:Gene: aq\_2059  
 C:Superfamily: transformation competence-related protein comF

Query Match 2.0%; Score 7; DB 2; Length 213;  
 Best Local Similarity 100.0%; Pred. No. 61;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 180 DLFLSE 186  
 |||||||  
 Db 9 DLFLSE 15

RESULT 35

T18061  
 hypothetical protein A559L - Chlorella virus PBCV-1

C:Species: Chlorella virus PBCV-1  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T18061  
 R:Graves, M.V.; Van Etten, J.L.  
 submitted to the EMBL Data Library, May 1999  
 A:Reference number: Z18806  
 A:Accession: T18061  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-213 <GRA>  
 A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC6918.1  
 A:Experimental source: specific host Chlorella strain NC64A  
 C:Genetics:  
 A:Note: A559L

Query Match 2.0%; Score 7; DB 2; Length 213;  
 Best Local Similarity 100.0%; Pred. No. 61;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 VLLAIVA 12  
 |||||||  
 Db 10 VLLAIVA 16

RESULT 36

E72401  
 conserved hypothetical protein - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima  
 C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
 C:Accession: E72401  
 R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hic  
 Garrell, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,  
 C.M.  
 Nature 399, 323-329, 1999  
 A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome  
 A:Reference number: A72200; MUID:99287316  
 A:Accession: E72401

A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-224 <ARN>  
 A:Cross-references: GB:AE001707; GB:AE000512; NID:g4980720; PIDN:AA035318.1; PID:g498072  
 A:Experimental source: strain MSB8  
 C:Genetics:  
 A:Gene: TM0226  
 C:Superfamily: conserved hypothetical protein b2966

Query Match 2.0%; Score 7; DB 2; Length 224;  
 Best Local Similarity 100.0%; Pred. No. 64;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LIALVAG 13  
 |||||  
 Db 61 LIALVAG 67

RESULT 37  
 A83413 molibdenum transport protein ModB PA1862 [Imported] - Pseudomonas aeruginosa (strain PAO  
 C:Species: Pseudomonas aeruginosa  
 C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: A83413  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bt  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
 .; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho  
 A:Reference number: A82950; MUID:20437337  
 A:Accession: A83413  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-228 <STO>  
 A:Cross-references: GB:AE004612; GB:AE004091; NID:g9947846; PIDN:AG05251.1; GSPDB:GN001  
 C:Genetics:  
 A:Experimental source: strain PAO1  
 A:Gene: modB; PA1862  
 C:Superfamily: maltose transport protein malG

Query Match 2.0%; Score 7; DB 2; Length 228;  
 Best Local Similarity 100.0%; Pred. No. 65;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FVLLAL 10  
 |||||  
 Db 211 FVLLAL 217

RESULT 38  
 Q0BCPE  
 tRNA nucleotidyltransferase (EC 2.7.7.56) - Escherichia coli  
 N:Alternate names: phosphate-dependent exoribonuclease; ribonuclease PH; RNase PH  
 C:Species: Escherichia coli  
 C>Date: 25-Feb-1985 #sequence\_revision 25-Feb-1985 #text\_change 18-Jun-1999  
 C:Accession: A04470; E65165  
 R:Poulsen, P.; Bonckamp, F.; Jensen, K.F.  
 EMBO J 3, 1783-1790, 1984  
 A:Title: Structure of the Escherichia coli pyre operon and control of pyre expression by  
 A:Reference number: A30400; MUID:85003588  
 A:Accession: A04470  
 A:Molecule type: DNA  
 A:Residues: 1-238 <POU>  
 A:Cross-references: GB:X00781; GB:X01713; NID:g499180; PIDN:CA25357.1; PID:g499181  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Ferna, N.T.; Burland, V.; Riley, M.; C  
 .A.; Rose, D.J.; Mau, B.; Sano, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:97426617  
 A:Accession: E65165  
 A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
 A:Residues: 1-122, 'GG', 125-223, 'ESNPL' <BLAT>  
 A:Cross-references: GB:AE000441; GB:U00096; NID:g1790063; PIDN:AGC76667.1; PID:g17900  
 A:Experimental source: strain K-12, substrain MG1655  
 A:Note: this strain is known to contain a frameshift mutation at the carboxyl end  
 R:Burland, V.; Plunkett III, G.; Daniels, D.L.; Blattner, F.R.  
 Genomics 16, 551-561, 1993  
 A:Title: DNA sequence and analysis of 136 kilobases of the Escherichia coli genome: o  
 A:Reference number: A45741; MUID:95315143  
 A:Contents: annotation; coding regions  
 R:Kelly, K.O.; Deutscher, M.P.  
 J. Biol. Chem. 267, 17153-17158, 1992  
 A:Title: Characterization of Escherichia coli RNase PH.  
 A:Reference number: A43128; MUID:92381029  
 A:Contents: annotation; enzyme activity  
 C:Genetics:  
 A:Gene: rph  
 A:Map position: 82 min  
 A:Note: this gene and the pyre gene constitute an operon  
 C:Function:  
 A:Description: catalyzes the reversible phosphorylation of ribonucleotides following -C  
 C:Superfamily: tRNA nucleotidyltransferase  
 C:Keywords: nucleotidyltransferase; tRNA processing  
 F:1-228/Product: tRNA nucleotidyltransferase #status predicted <MAT>

Query Match 2.0%; Score 7; DB 1; Length 238;  
 Best Local Similarity 100.0%; Pred. No. 67;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 240 LVDAOK 246  
 |||||  
 Db 137 LVDAOK 143

RESULT 39  
 B84249  
 hypothetical protein Vng0927c [Imported] - Halobacterium sp. NRC-1  
 C:Species: Halobacterium sp. NRC-1  
 C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: B84249  
 R:Ng, W.V.; Kennedy, S.P.; Mahrtas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky  
 ; Leitner, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja  
 Jung, K.H.; Alam, M.; Freitas, T.  
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.;  
 A:Title: Genome sequence of Halobacterium species NRC-1.  
 A:Reference number: A84160; MUID:20504483  
 A:Accession: B84249  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-239 <STO>  
 A:Cross-references: GB:AE004437; NID:g10580486; PIDN:AG19358.1; GSPDB:GN00138  
 C:Genetics:  
 A:Gene: VNG0927C

Query Match 2.0%; Score 7; DB 2; Length 239;  
 Best Local Similarity 100.0%; Pred. No. 68;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VEVVLLA 9  
 |||||  
 Db 215 VEVVLLA 221

RESULT 40  
 T38152  
 conserved hypothetical protein SPAC22A12.12c [Imported] - fission yeast (Schizosaccha  
 C:Species: Schizosaccharomyces pombe  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 28-Jul-2000  
 C:Accession: T38152  
 R:Pearson, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.

submitted to the EMBL Data Library, September 1997  
A:Reference number: Z21774

A:Accession: T38152

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-240 <PEP>

A:Cross-references: EMBL:Z29295; PIDN:CAB16582.1; GSPDB:GN00066; SPDB:SPAC22A12.12C

A:Experimental source: strain 972h-; cosmid c22A12

A:Genetic: SPDB:SPAC22A12.12C

A:Map position: 1

Query Match 2.0%; Score 7; DB 2; Length 240;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 38 PGERIPD 44  
|||||||

DB 15 PGERIPD 21

#### RESULT 41

E71305

conserved hypothetical protein TP0588 - syphilis spirochete

C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 05-Nov-1999

C:Accession: E71305

R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwilt

ron, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo

lson, L.; Weidman, J.; Smith, H.O.; Yenter, J.C.

Science 281, 375-388, 1998

A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A:Reference number: A71250; MUID:98332770

A:Accession: E71305

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-254 <COD>

A:Cross-references: GB:AE001234; GB:AE000520; NID:93322881; PIDN:MAC65563.1; PID:9332288

A:Experimental source: strain Nichols

C:Genetics:

A:Gene: TP0588

Query Match 2.0%; Score 7; DB 2; Length 254;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 191 HDISL 197  
|||||||

DB 181 HDISL 187

#### RESULT 42

D75404 phosphatidylylglycerophosphatase B-related protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000

C:Accession: D75404

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

S.; Smith, H.O.; Yenter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896

A:Accession: D75404

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-260 <WH>

A:Cross-references: GB:AE001982; GB:AE000513; NID:96459109; PIDN:AAF10927.1; PID:9645911

A:Experimental source: strain R1

C:Genetics:

A:Gene: DRL355

A:Map position: 1

Query Match 2.0%; Score 7; DB 2; Length 260;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 LALVAGV 14  
|||||||

DB 87 LALVAGV 93

#### RESULT 43

T39301

probable o-methyltransferase - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jul-2000

C:Accession: T39301

R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.

submitted to the EMBL Data Library, March 1998

A:Reference number: Z21843

A:Accession: T39301

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-266 <MOO>

A:Cross-references: EMBL:AL022117; PIDN:CAI17918.1; GSPDB:GN00067; SPDB:SPB119.03

A:Experimental source: strain 972h-; cosmid c119

C:Genetics:

A:Gene: SPDB:SPB119.03

A:Map position: 2

C:Superfamily: caffeine-1-COA 3-O-methyltransferase

Query Match 2.0%; Score 7; DB 2; Length 266;  
Best Local Similarity 100.0%; Pred. No. 74;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 216 LAGDEI 222  
|||||||

DB 121 LAGDEI 127

#### RESULT 44

G69791

hypothetical protein yebC - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000

C:Accession: G69791

R:Kunst, F.; Onasawa, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber

C.; Bron, S.; Brouillette, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal

lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hull, M

koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lard

Y.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau

Riger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanl

A:Authors: Schreier, R.; Schroeder, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Se

akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Toyonari, A.; Tosato, V.; Uchiya

T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Zamamoto, H.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis

A:Reference number: A69580; MUID:98044033

A:Accession: G69791

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-267 <KUN>

A:Cross-references: GB:Z99107; GB:AL009126; NID:92632866; PIDN:CAB12457.1; PID:9263229

A:Experimental source: strain 168

C:Genetics:

A:Gene: yebC

C:Superfamily: Bacillus subtilis hypothetical protein yebC

Query Match  
Best Local Similarity 100.0%; Score 7; DB 2; Length 267;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 F51IKSP 25  
|||||||  
Db 75 F51IKSP 81

## RESULT 45

E83215  
Probable permease of ABC transporter PA3448 [Imported] - *Pseudomonas aeruginosa* (strain C:Species: *Pseudomonas aeruginosa*  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: E83215  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Berman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapid, K.; Lim, J.; Lory, S.; Olson, M.V.  
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen  
A:Reference number: A82950; MUID:20437337  
A:Accession: E83215  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-274 <STO>  
A:Cross-references: GB:AE004765; GB:AE004091; NID:9949580; PIDN:AG06836.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA3448  
C:Superfamily: Synechococcus nitrate transport protein nrtb

Query Match  
Best Local Similarity 100.0%; Score 7; DB 2; Length 274;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 LAIVAGY 14  
|||||||  
Db 153 LAIVAGY 159

## RESULT 46

E83909  
Oligopeptide ABC transporter (ATP-binding protein) BH2077 [Imported] - *Bacillus halodurans*  
C:Species: *Bacillus halodurans*  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 31-Dec-2000  
C:Accession: E83909  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and  
A:Reference number: A83650; MUID:20263314  
A:Accession: E83909  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-274 <STO>  
A:Cross-references: GB:AP001514; GB:BA000004; NID:q10174613; PIDN:BA005796.1; GSPDB:GN00  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BH2077

Query Match  
Best Local Similarity 100.0%; Score 7; DB 2; Length 274;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 313 IMIALAL 319  
|||||||  
Db 159 IMIALAL 165

## RESULT 47

E75188  
morphine 6-dehydrogenase homolog PAB2329 - *Pyrococcus abyssi* (strain Orsay)  
C:Species: *Pyrococcus abyssi*  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C:Accession: F75188  
R:anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999

A:Description: *Pyrococcus abyssi* genome sequence: insights into archaeal chromosome s  
A:Reference number: A75001  
A:Accession: F75188  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-279 <KAM>  
A:Cross-references: GB:A1248283; GB:AL096836; NID:95457433; PIDN:CA848957.1; PID:9545  
C:Genetics:  
A:Gene: PAB2329  
C:Superfamily: aldehyde reductase

Query Match  
Best Local Similarity 100.0%; Score 7; DB 2; Length 279;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 221 EIGKRYG 227  
|||||||  
Db 221 EIGKRYG 227

## RESULT 48

T00880  
hypothetical protein At2g45670 [Imported] - *Arabidopsis thaliana*  
N:Alternate names: hypothetical protein F17K2.20  
C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 16-Feb-2001  
C:Accession: T00880; D84893  
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Syke  
submitted to the EMBL Data Library, March 1998  
A:Description: *Arabidopsis thaliana* chromosome II BAC F17K2 genomic sequence.  
A:Reference number: Z14207  
A:Accession: T00880  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-281 <ROU>

A:Cross-references: EMBL:AC003680; NID:92979540; PID:92979560  
A:Experimental source: cultivar Columbia  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanden, S.E.; Umeyam, L.; Tallon,  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.  
A:Reference number: A84420; MUID:20083487  
A:Accession: D84893  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-281 <STO>  
A:Cross-references: GB:AE002093; NID:92979560; PIDN:AA006169.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: AT2g45670; F17K2.20  
A:Map position: 2  
A:Insertions: 155/2; 215/3; 235/3  
C:Superfamily: *Arabidopsis thaliana* hypothetical protein F17K2.20

Query Match  
Best Local Similarity 100.0%; Score 7; DB 2; Length 281;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 162 NSVLSL 168  
|||||||  
Db 269 NSVLSL 275

RESULT 49

B70928  
 Probable dact protein - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
 C:Accession: B70928  
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Ruter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70500; M0ID:98295987  
 A:Accession: B70928  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-291 <COL>  
 A:Cross-references: GB:Z74024; GB:AL123456; NID:g3250700; PIDN:CA98348.1; PID:e248782;  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: dactb

Query Match 2.0%; Score 7; DB 2; Length 291;  
 Best Local Similarity 100.0%; Pred. No. 80;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 6 VLALVA 12  
 |||||  
 Db 73 VLALVA 79

RESULT 50  
 S76631  
 hypothetical protein - Synechocystis sp. (strain PCC 6803)  
 C:Species: Synechocystis sp.  
 A:Variety: PCC 6803  
 C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
 C:Accession: S76631  
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O., K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
 DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
 S.  
 A:Reference number: S74322; M0ID:97061201  
 A:Accession: S76631  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-299 <KAN>  
 A:Cross-references: EMBL:D64004; GB:AB001339; NID:g1001701; PIDN:BA10575.1; PID:d101122  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 2.0%; Score 7; DB 2; Length 299;  
 Best Local Similarity 100.0%; Pred. No. 82;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 182 LFLSELQ 188  
 |||||  
 Db 66 LFLSELQ 72

Search completed: August 17, 2001, 16:46:15  
 Job time: 137 sec





85	6	1.7	20	16	AA886668	Human glutamic ac
86	6	1.7	20	16	AA872281	Glutamic acid deca
87	6	1.7	20	17	AAW16993	plgC SH3 domain-bl
88	6	1.7	20	18	AAW18849	65 kD glutamic aci
89	6	1.7	20	21	AA822979	White shrimp multi
90	6	1.7	20	21	AAW59554	GAD65 fragment, pe
91	6	1.7	21	20	AAW23233	peptide derived fr
92	6	1.7	21	21	AA822980	White shrimp hydro
93	6	1.7	24	14	AA836918	Insulin-like growt
94	6	1.7	24	21	AA822988	White shrimp hydro
95	6	1.7	24	22	AA866904	Maize hydroxyroll
96	6	1.7	25	8	AAW71150	Sequence of peptid
97	6	1.7	25	21	AA822987	White shrimp hydro
98	6	1.7	27	8	AAW71149	Sequence of peptid
99	6	1.7	27	20	AAW11724	Sequence of peptid
100	6	1.7	27	20	AAW12041	Human 5' EST seque
101	6	1.7	27	21	AA824567	Human secreted pro
102	6	1.7	29	8	AAW71146	Sequence of peptid
103	6	1.7	29	8	AAW71147	Human 5' EST relat
104	6	1.7	29	21	AAW65388	New Insulin-like g
105	6	1.7	30	10	AAW61464	CDR3 sequence of T
106	6	1.7	30	16	AAW74845	CDR3 sequence of T
107	6	1.7	31	16	AAW33840	Peptide RIV from t
108	6	1.7	32	21	AAW30487	Amino acid sequence
109	6	1.7	33	7	AAW60717	Synthetic sequence
110	6	1.7	34	20	AAW33058	Human secreted pro
111	6	1.7	35	9	AAW80431	Preprosequence of
112	6	1.7	38	13	AAW75584	Diguettia toxin DK
113	6	1.7	38	16	AAW80152	Diguettia spider ve
114	6	1.7	40	18	AAW4276	Rat cadherin-9 cyt
115	6	1.7	40	18	AAW25650	Arabidopsis thalia
116	6	1.7	41	21	AAW31115	Sequence of region
117	6	1.7	44	14	AAW38140	Human 5' EST seque
118	6	1.7	44	20	AAW38216	C-phycocyanin-1a
119	6	1.7	45	11	AAW12865	Bovine IgT1 parlia
120	6	1.7	45	11	AAW12865	Dog IgT1 parlia s
121	6	1.7	46	21	AAW06157	Characteristic pro
122	6	1.7	48	15	AAW54008	Polyketide synthas
123	6	1.7	52	21	AAW82098	Del-1 epidermal gr
124	6	1.7	57	20	AAW94686	M. bovis Dalton 2d
125	6	1.7	59	22	AAW26613	Human src-family k
126	6	1.7	62	10	AAW91501	New insulin-like g
127	6	1.7	63	21	AAW06627	Derivative of Insu
128	6	1.7	64	21	AAW12929	Human secreted pro
129	6	1.7	65	14	AAW36768	Arabidopsis thalia
130	6	1.7	65	20	AAW39198	Sequence of a pept
131	6	1.7	65	21	AAW36035	Extended human sec
132	6	1.7	67	14	AAW22566	Zea mays protein f
133	6	1.7	67	17	AAW36847	Insulin-like growt
134	6	1.7	67	20	AAW89753	O-glycosylated, tr
135	6	1.7	67	21	AAW41696	Human PRO617 (UNQ3
136	6	1.7	67	21	AAW44252	Human PRO617 used
137	6	1.7	67	21	AAW19581	A. thaliana TINY E
138	6	1.7	68	20	AAW31476	Zea mays protein f
139	6	1.7	68	21	AAW25835	Arabidopsis thalia
140	6	1.7	68	21	AAW10695	Arabidopsis thalia
141	6	1.7	68	21	AAW37807	O-glycosylated, tr
142	6	1.7	69	17	AAW89754	Arabidopsis thalia
143	6	1.7	69	17	AAW23745	Arabidopsis thalia
144	6	1.7	69	21	AAW51168	Seq ID 2 used in t
145	6	1.7	70	5	AAW40034	Sequence of human
146	6	1.7	70	8	AAW70414	Sequence of oxidat
147	6	1.7	70	10	AAW1539	New insulin-like g
148	6	1.7	70	10	AAW91502	Analogous IGF130 of
149	6	1.7	70	10	AAW94661	Analogous IGF132 of
150	6	1.7	70	10	AAW94660	Human insulin-like
151	6	1.7	70	11	AAW06307	
152	6	1.7	70	11	AAW06307	
153	6	1.7	70	11	AAW06307	
154	6	1.7	70	11	AAW06307	
155	6	1.7	70	11	AAW06307	
156	6	1.7	70	11	AAW06307	
157	6	1.7	70	11	AAW06307	



231	6	1.7	94	16	AAR80153	Diguettia toxin DK
232	6	1.7	94	18	AAM14275	Diguettia spider ve
233	6	1.7	94	18	AAM32104	Diguettia spp. spid
234	6	1.7	94	21	AA607749	Arabidopsis thalia
235	6	1.7	94	21	AA607749	Arabidopsis thalia
236	6	1.7	95	14	AA637549	Sequence of insuli
237	6	1.7	95	19	AA667492	Lamb signal sequen
238	6	1.7	96	21	AA619133	Zea mays protein f
239	6	1.7	99	21	AA619133	Zea mays protein f
240	6	1.7	101	9	AA682123	Human tropotin 1 p
241	6	1.7	101	20	AA676598	Fusion protein of
242	6	1.7	102	18	AA610370	Human ovarian tumo
243	6	1.7	102	21	AA625465	Human Del-1 splice
244	6	1.7	104	21	AA619058	Pinus radiata cell
245	6	1.7	105	21	AA653419	Zea mays protein f
246	6	1.7	106	20	AA653535	Human colon cancer
247	6	1.7	106	21	AA653582	Chlamydia pneumonia
248	6	1.7	107	21	AA608514	Arabidopsis thalia
249	6	1.7	107	21	AA652702	Arabidopsis thalia
250	6	1.7	108	21	AA651932	Arabidopsis thalia
251	6	1.7	108	21	AA655981	Zea mays protein f
252	6	1.7	109	21	AA676074	Arabidopsis thalia
253	6	1.7	109	22	AA645738	Human skin cell pr
254	6	1.7	109	22	AA656013	Human BNP propepti
255	6	1.7	109	22	AA656013	Skin cell protein
256	6	1.7	110	20	AA659705	Secreted protein 5
257	6	1.7	110	21	AA601697	Human secreted pro
258	6	1.7	110	21	AA678227	Human signal pepti
259	6	1.7	110	21	AA678227	Human signal pepti
260	6	1.7	112	20	AA632774	HTV SF-2 strain ch
261	6	1.7	112	21	AA659044	Arabidopsis thalia
262	6	1.7	115	20	AA643624	Human 1-40 OH grow
263	6	1.7	115	21	AA603796	A portion of the s
264	6	1.7	115	21	AA603796	Human secreted pro
265	6	1.7	116	21	AA678250	Human signal pepti
266	6	1.7	117	21	AA655580	Human secreted pro
267	6	1.7	117	21	AA655580	Arabidopsis thalia
268	6	1.7	118	18	AA609772	Human secreted pro
269	6	1.7	118	20	AA612411	Killer toxin leade
270	6	1.7	119	7	AA605783	Human 5' EST secre
271	6	1.7	119	21	AA695693	Human prepro-somat
272	6	1.7	120	20	AA697368	Cosmid cHR15 enco
273	6	1.7	120	21	AA602003	Murine growth horm
274	6	1.7	121	18	AA623301	Human secreted pro
275	6	1.7	122	21	AA617769	Rabbit insulin Iik
276	6	1.7	122	21	AA649720	Arabidopsis thalia
277	6	1.7	123	21	AA656934	Arabidopsis thalia
278	6	1.7	124	20	AA697371	Human prostate can
279	6	1.7	125	17	AA688459	Murine/porcine chi
280	6	1.7	125	20	AA697365	erBB-3 fragment.
281	6	1.7	125	20	AA697366	Porcine growth horm
282	6	1.7	125	20	AA697367	Porcine growth horm
283	6	1.7	125	20	AA697369	Ovine growth hormo
284	6	1.7	125	21	AA638339	Caprine growth hor
285	6	1.7	129	20	AA638339	Arabidopsis thalia
286	6	1.7	129	21	AA638339	Neisseria meningit
287	6	1.7	129	21	AA655987	Arabidopsis thalia
288	6	1.7	129	21	AA655987	Arabidopsis thalia
289	6	1.7	130	19	AA642029	Clone L68_2 protei
290	6	1.7	130	20	AA608637	Murine secreted pr
291	6	1.7	130	21	AA673328	Murine secreted pr
292	6	1.7	131	15	AA683193	hEGF-ST-IGF-1 fusl
293	6	1.7	133	6	AA650927	Sequence of human
294	6	1.7	133	21	AA618737	Zea mays protein f
295	6	1.7	134	11	AA604087	Protein encoded by
296	6	1.7	134	11	AA605603	Human Brain Natriu
297	6	1.7	134	20	AA605603	Human gamma-BNP pr
298	6	1.7	134	21	AA632548	Eucalyptus grandis
299	6	1.7	135	22	AA645735	Human BNP preprope
300	6	1.7	136	21	AA652442	Mycobacterium tube
301	6	1.7	137	6	AA656824	Arabidopsis thalia
302	6	1.7	137	8	AA670378	Sequence of human
303	6	1.7	137	8	AA670378	proctected human in
					AA670101	Sequence encoded b

Protecting peptide  
Pinus radiata cell  
Arabidopsis thalia  
Sequence of human  
Zea mays protein f  
Baboon reovirus fu  
Arabidopsis thalia  
Human secreted pro  
Streptococcus pneu  
Human ORF ORF15  
Human Hydrilase pr  
Amino acid sequenc  
Arabidopsis thalia  
Puroindoline A pro  
Serine substituted  
Arabidopsis thalia  
Sequence of a pept  
Human secreted pro  
Arabidopsis thalia  
Insulin-like growth  
Human IGF-1. Homo  
Human IGF-1 protei  
Yeast alpha-mating  
Goat Insulin Like  
Arabidopsis thalia  
Yeast alpha-factor  
Chimeric rhIGF-I-A  
Human insulin like  
H. pylori membrane  
H. pylori transmem  
Human ORF ORF213  
Staphylokinase-som  
Amino acid sequenc  
Human secreted pro  
Fragment of human  
Human pancreatic c  
Human cell cycle a  
Human secreted pro  
S. pneumoniae pfs  
Parvo virus B19 PA  
Arabidopsis thalia  
Arabidopsis thalia  
Sequence of a pept  
Human secreted pro  
Arabidopsis thalia  
C-terminal domain  
C-terminal domain  
C-terminal domain  
Parvo virus B19 PA  
Human ORF ORF1025  
Human secreted pro  
Protein encoded by  
Gene 18 human secr  
C. glutamicum SRT  
Corynebacterium gl  
Arabidopsis thalia  
Arabidopsis thalia  
Human secreted pro  
A krill derived pr  
Human colon cancer  
White shrimp multi  
White shrimp multi  
Amino acid sequenc  
S. pneumoniae N-ce  
Human adenovirus t  
Arabidopsis thalia  
A human prostate s  
Hydroxyphenyl pyru



523	6	1.7	271	21	AAAG33382	Arabidopsis thalia	596	6	1.7	314	21	AAAG37868	Arabidopsis thalia
524	6	1.7	272	21	AAAB58955	Breast and ovarian	597	6	1.7	314	21	AAAV70572	Mouse Phox2b prote
525	6	1.7	272	21	AAAG20510	Arabidopsis thalia	598	6	1.7	316	21	AAAV95690	Cosmid CHR1M5 enco
526	6	1.7	272	21	AAAG35062	Arabidopsis thalia	599	6	1.7	316	21	AAAV95734	Cosmid CHR1M5 enco
527	6	1.7	276	21	AAAG27473	Arabidopsis thalia	600	6	1.7	318	19	AAAV52285	Gallus domesticus
528	6	1.7	276	21	AAAG548497	Arabidopsis thalia	601	6	1.7	318	19	AAAG22348	Arabidopsis thalia
529	6	1.7	276	21	AAAG56474	Arabidopsis thalia	602	6	1.7	319	19	AAAM41572	Human cardiac trop
530	6	1.7	277	19	AAAW71514	Helicobacter polyp	603	6	1.7	319	21	AAAG27869	Human cardiac trop
531	6	1.7	277	21	AAAB25312	Eucalyptus grandis	604	6	1.7	320	20	AAAB88326	Arabidopsis thalia
532	6	1.7	277	21	AAAG10705	Arabidopsis thalia	605	6	1.7	320	20	AAAG39641	Abegusyl transfer
533	6	1.7	278	17	AAAM01416	Arabidopsis thalia	606	6	1.7	321	21	AAAG24742	Arabidopsis thalia
534	6	1.7	281	21	AAAG21219	H. pylori GHP0 721	607	6	1.7	321	21	AAAG45526	Arabidopsis thalia
535	6	1.7	282	19	AAAG98705	Arabidopsis thalia	608	6	1.7	323	19	AAAM53246	Human haematopoiet
536	6	1.7	283	21	AAAG56473	Arabidopsis thalia	609	6	1.7	323	20	AAAY14523	Human lymphocytic
537	6	1.7	283	21	AAAG59189	Arabidopsis thalia	610	6	1.7	324	21	AAAV75253	Neisseria meningit
538	6	1.7	283	21	AAAG59435	Arabidopsis thalia	611	6	1.7	324	21	AAAV75254	Neisseria meningit
539	6	1.7	285	21	AAAG12202	Arabidopsis thalia	612	6	1.7	324	21	AAAV75255	Neisseria meningit
540	6	1.7	285	21	AAAG47575	Arabidopsis thalia	613	6	1.7	325	21	AAAG13668	Arabidopsis thalia
541	6	1.7	285	21	AAV76141	Human secreted pro	614	6	1.7	327	21	AAAG50174	Arabidopsis thalia
542	6	1.7	287	21	AAAB04181	Leukocyte immunogl	615	6	1.7	328	19	AAAM64776	Floral organ-speci
543	6	1.7	287	21	AAAB35708	Human protein cont	616	6	1.7	328	21	AAAB25215	Eucalyptus grandis
544	6	1.7	287	21	AAAG31225	Arabidopsis thalia	617	6	1.7	328	21	AAAG24741	Arabidopsis thalia
545	6	1.7	289	11	AAAR07523	Alpha-Trichosanthe	618	6	1.7	330	21	AAAG22347	Arabidopsis thalia
546	6	1.7	289	11	AAAR07514	Trichosanthe from	619	6	1.7	331	21	AAAG17979	Arabidopsis thalia
547	6	1.7	289	13	AAAR25572	Trichosanthe from	620	6	1.7	331	21	AAAG27868	Arabidopsis thalia
548	6	1.7	289	13	AAAR29276	Ribosome inactivat	621	6	1.7	331	21	AAAG45525	Arabidopsis thalia
549	6	1.7	289	13	AAAR29272	Trichosanthe prot	622	6	1.7	331	21	AAAG45539	Leucocytosoma prot
550	6	1.7	289	14	AAAR32986	Encodes chinese cu	623	6	1.7	332	16	AAAR70492	Human GABA recept
551	6	1.7	289	15	AAAR5129	Alpha-trichosanthe	624	6	1.7	332	20	AAAY34112	Human ORF1354
552	6	1.7	289	15	AAAM10468	Chinese cucumber a	625	6	1.7	332	20	AAAY30312	Human ORF1354
553	6	1.7	289	18	AAAM11870	Chinese cucumber a	626	6	1.7	332	20	AAAY29662	Human 7TM receptor
554	6	1.7	289	20	AAV01374	Chinese cucumber a	627	6	1.7	332	21	AAAG44088	Arabidopsis thalia
555	6	1.7	289	20	AAAM84192	Chinese cucumber a	628	6	1.7	334	21	AAAG20508	Arabidopsis thalia
556	6	1.7	289	21	AAV87791	Chinese cucumber a	629	6	1.7	334	21	AAAG35060	Arabidopsis thalia
557	6	1.7	290	21	AAAG23235	Arabidopsis thalia	630	6	1.7	335	21	AAAB52610	Helicobacter pylor
558	6	1.7	290	21	AAAG34779	Arabidopsis thalia	631	6	1.7	335	20	AAV37249	Protein which is s
559	6	1.7	290	21	AAAG47574	Arabidopsis thalia	632	6	1.7	336	21	AAAG07750	Arabidopsis thalia
560	6	1.7	291	21	AAAG12201	Arabidopsis thalia	633	6	1.7	336	21	AAAG40592	Arabidopsis thalia
561	6	1.7	291	21	AAAG13753	Arabidopsis thalia	634	6	1.7	337	21	AAAB41590	Human ORF1354
562	6	1.7	291	21	AAAG1750	Arabidopsis thalia	635	6	1.7	337	21	AAV77845	Human serine/threo
563	6	1.7	294	21	AAAG21218	Arabidopsis thalia	636	6	1.7	340	21	AAAG33319	Zea mays protein f
564	6	1.7	294	21	AAAG42189	Arabidopsis thalia	637	6	1.7	340	21	AAAG45958	Arabidopsis thalia
565	6	1.7	294	21	AAAG44089	Arabidopsis thalia	638	6	1.7	341	15	AAAR59522	GAD65 1-24 N-term
566	6	1.7	295	22	AAAB79664	Corynebacterium g1	639	6	1.7	341	19	AAAM69760	Acetobacter xylinu
567	6	1.7	295	22	AAAB76677	Corynebacterium g1	640	6	1.7	344	16	AAAR75359	B. campestris clon
568	6	1.7	296	18	AAAM20802	H. pylori inner me	641	6	1.7	344	18	AAAM15417	RA9 protein. Bras
569	6	1.7	296	21	AAAG07338	Arabidopsis thalia	642	6	1.7	344	19	AAAM30528	Tomato fruit-speci
570	6	1.7	297	21	AAAG33320	Zea mays protein f	643	6	1.7	344	19	AAAM59610	Brassica campestr
571	6	1.7	297	21	AAAG45527	Arabidopsis thalia	644	6	1.7	344	19	AAAM30672	Brassica campestr
572	6	1.7	299	21	AAAB04180	Leukocyte immunogl	645	6	1.7	344	20	AAAB87714	Brassica campestr
573	6	1.7	300	19	AAAM50195	A krill derived pr	646	6	1.7	344	21	AAAB12926	An active acyltran
574	6	1.7	300	19	AAAM50197	A krill derived pr	647	6	1.7	344	21	AAAG05574	Protein sequence o
575	6	1.7	300	20	AAAV35629	Chlamydia pneumoni	648	6	1.7	344	21	AAAG37001	Arabidopsis thalia
576	6	1.7	300	21	AAAB22947	White shrimp multi	649	6	1.7	344	21	AAV49928	Brassica campestr
577	6	1.7	300	21	AAAV93940	Amino acid sequenc	650	6	1.7	345	5	AAV40674	Sequence encoded b
578	6	1.7	301	20	AAAV34825	Chlamydia pneumoni	651	6	1.7	345	6	AAV50873	Methionyl-synthet
579	6	1.7	301	21	AAV57945	Human transmembran	652	6	1.7	345	21	AAAG50173	Arabidopsis thalia
580	6	1.7	302	19	AAAM50200	A krill derived pr	653	6	1.7	346	21	AAAB53523	Helicobacter pylor
581	6	1.7	302	21	AAAB22951	White shrimp multi	654	6	1.7	346	21	AAV68647	Amino acid sequenc
582	6	1.7	302	21	AAV93945	Amino acid sequenc	655	6	1.7	346	22	AAAB31732	Mastadenovirus 23
583	6	1.7	304	21	AAAG20509	Arabidopsis thalia	656	6	1.7	346	22	AAAB67175	Human secreted pro
584	6	1.7	304	21	AAAG35061	Arabidopsis thalia	657	6	1.7	346	22	AAAB27564	Human secreted pro
585	6	1.7	305	21	AAAG24743	Arabidopsis thalia	658	6	1.7	349	17	AAAR79443	Galanin receptor.
586	6	1.7	306	21	AAAG19283	Arabidopsis thalia	659	6	1.7	349	17	AAAR95070	Human galanin rece
587	6	1.7	307	21	AAAG37869	Arabidopsis thalia	660	6	1.7	349	21	AAAG20012	Arabidopsis thalia
588	6	1.7	308	19	AAAG50198	A krill derived mu	661	6	1.7	349	21	AAAG25349	Arabidopsis thalia
589	6	1.7	308	21	AAAG56472	Arabidopsis thalia	662	6	1.7	351	20	AAAM88296	Wheat geminivirus
590	6	1.7	308	21	AAAG58188	Arabidopsis thalia	663	6	1.7	351	21	AAAG12878	Arabidopsis thalia
591	6	1.7	308	21	AAAG59434	Arabidopsis thalia	664	6	1.7	351	21	AAAG42046	Arabidopsis thalia
592	6	1.7	310	21	AAAG48496	Arabidopsis thalia	665	6	1.7	351	21	AAAG53633	Arabidopsis thalia
593	6	1.7	311	21	AAAG22261	Arabidopsis thalia	666	6	1.7	351	21	AAV84439	Amino acid sequenc
594	6	1.7	313	19	AAAM55983	Swinepox virus str	667	6	1.7	351	22	AAAB69050	Human ribonucleoti
595	6	1.7	314	18	AAAM14283	Human neuroblastom	668	6	1.7	352	20	AAAM93953	Human regulatory m

669	6	1.7	352	21	AAAG10704	Arabidopsis thalia	742	6	1.7	391	21	AAV68644	Amino acid sequenc
670	6	1.7	352	21	AAAG53618	Arabidopsis thalia	743	6	1.7	391	21	AAV68655	Amino acid sequenc
671	6	1.7	353	19	AAAM61370	Human H1CB splice	744	6	1.7	391	22	AAAB31728	Amino acid sequenc
672	6	1.7	353	20	AAV16570	Amino acid sequenc	745	6	1.7	391	22	AAAB31740	Amino acid sequenc
673	6	1.7	353	21	AAV97310	Murine H1CB recep	746	6	1.7	391	22	AAAB67172	Mastadenovirus 13
674	6	1.7	353	21	AAAB12778	Rat SLC-1 protein	747	6	1.7	391	22	AAAB67183	Mastadenovirus 33
675	6	1.7	353	21	AAAB13437	Rat MCH1 receptor	748	6	1.7	392	14	AAAB43074	Aspergillus niger
676	6	1.7	353	21	AAAB13442	Truncated human MC	749	6	1.7	392	21	AAAB12137	Hydrophobic domain
677	6	1.7	353	21	AAAG13752	Arabidopsis thalia	750	6	1.7	392	21	AAV53043	Human secreted pro
678	6	1.7	353	21	AAAG51749	Arabidopsis thalia	751	6	1.7	393	18	AAAB32867	Mitogen activated
679	6	1.7	353	21	AAV90258	Human H1CB protei	752	6	1.7	393	19	AAAB37149	Murine Ena-Vasp 11
680	6	1.7	353	21	AAV57284	Human GPCR protein	753	6	1.7	393	21	AAV94478	Human MEX1 protein
681	6	1.7	353	22	AAV97668	Human MCH-R1 prote	754	6	1.7	394	21	AAAB43943	Human cancer assoc
682	6	1.7	353	22	AAAB48152	Rat melamin-concen	755	6	1.7	394	21	AAV68669	Amino acid sequenc
683	6	1.7	354	21	AAAG42045	Arabidopsis thalia	756	6	1.7	394	22	AAAB31754	Amino acid sequenc
684	6	1.7	354	21	AAAG44087	Arabidopsis thalia	757	6	1.7	394	22	AAAB67197	Mastadenovirus 49
685	6	1.7	355	21	AAAG17978	Arabidopsis thalia	758	6	1.7	395	21	AAAG09249	Arabidopsis thalia
686	6	1.7	355	21	AAAG45538	Arabidopsis thalia	759	6	1.7	395	21	AAAG42022	Arabidopsis thalia
687	6	1.7	356	21	AAAG10703	Arabidopsis thalia	760	6	1.7	395	22	AAAB79126	Arabidopsis thalia
688	6	1.7	357	21	AAAG13667	Arabidopsis thalia	761	6	1.7	395	22	AAAB79127	Corynebacterium gl
689	6	1.7	358	18	AAAM08135	Human cytokine res	762	6	1.7	396	12	AAAB13720	Corynebacterium gl
690	6	1.7	358	19	AAAM44246	Human HP4 proteagl	763	6	1.7	396	19	AAV55070	Human mevalonate K
691	6	1.7	358	21	AAV87954	Human CR3 protein	764	6	1.7	396	21	AAV96808	Streptococcus pneu
692	6	1.7	359	20	AAV37209	Protein which is s	765	6	1.7	397	21	AAAG64442	Streptococcus agal
693	6	1.7	361	21	AAAG52779	Arabidopsis thalia	766	6	1.7	400	19	AAV71487	Arabidopsis thalia
694	6	1.7	361	21	AAV71110	Human Hydrolase pr	767	6	1.7	400	20	AAV41652	Helicobacter polyp
695	6	1.7	362	22	AAAB36873	Rabbit leukocyte a	768	6	1.7	400	20	AAAB8434	Human MEX2 protein
696	6	1.7	362	19	AAAM80701	S. pneumoniae cati	769	6	1.7	400	21	AAV82775	Disease associated
697	6	1.7	362	20	AAV38750	Neisseria gonorrh	770	6	1.7	400	21	AAV44217	Human chordin rela
698	6	1.7	362	21	AAV91646	Human secreted pro	771	6	1.7	400	21	AAV53033	Soybean phytoene s
699	6	1.7	365	21	AAAG17977	Arabidopsis thalia	772	6	1.7	401	18	AAW22779	Human secreted pro
700	6	1.7	365	21	AAAG34778	Arabidopsis thalia	773	6	1.7	401	19	AAW38548	Human septin-2 pro
701	6	1.7	365	21	AAAG45537	Arabidopsis thalia	774	6	1.7	402	17	AAV98358	Streptococcus pneu
702	6	1.7	366	22	AAAB78930	C. glutamicum SRT	775	6	1.7	402	19	AAAB69339	Somatosulin-like
703	6	1.7	366	22	AAAB64409	Amino acid sequenc	776	6	1.7	402	21	AAV90623	Secreted protein o
704	6	1.7	367	21	AAAG2861	Arabidopsis thalia	777	6	1.7	402	21	AAV90657	Human G protein-co
705	6	1.7	367	21	AAV87207	Human secreted pro	778	6	1.7	402	22	AAAB46151	Human mutant G pro
706	6	1.7	368	18	AAAB32868	Mitogen activated	779	6	1.7	403	20	AAV34704	Human melanin conc
707	6	1.7	369	21	AAAG34777	Arabidopsis thalia	780	6	1.7	403	21	AAAG30833	Chlamydia pneumoni
708	6	1.7	370	19	AAAM62597	Human 7-transmembr	781	6	1.7	404	21	AAAG14701	Arabidopsis thalia
709	6	1.7	371	20	AAV41310	Human secreted pro	782	6	1.7	405	21	AAAG11844	Arabidopsis thalia
710	6	1.7	371	20	AAV38749	Neisseria meningit	783	6	1.7	405	21	AAAG53248	Arabidopsis thalia
711	6	1.7	371	20	AAV38751	Neisseria gonorrh	784	6	1.7	406	20	AAV02687	Human secreted pro
712	6	1.7	371	20	AAV38748	Neisseria meningit	785	6	1.7	410	19	AAAB6414	Human secreted pro
713	6	1.7	371	20	AAV08963	A. gossypii ORF 2	786	6	1.7	410	20	AAV93254	Mycobacterium bovi
714	6	1.7	371	21	AAAB43653	Human cancer assoc	787	6	1.7	411	20	AAV95259	Human ESRP1 protei
715	6	1.7	371	21	AAAG42860	Arabidopsis thalia	788	6	1.7	411	21	AAV95599	S. pneumoniae amps
716	6	1.7	372	19	AAAM41571	Cardiac troponin I	789	6	1.7	412	21	AAAG4700	Arabidopsis thalia
717	6	1.7	373	21	AAAG23381	Arabidopsis thalia	790	6	1.7	412	21	AAAG5811	Arabidopsis thalia
718	6	1.7	374	21	AAV68651	Amino acid sequenc	791	6	1.7	413	17	AAV95247	Arabidopsis thalia
719	6	1.7	374	22	AAAB31736	Amino acid sequenc	792	6	1.7	413	19	AAAB48390	Human secreted pro
720	6	1.7	374	22	AAAB67179	Mastadenovirus 28	793	6	1.7	413	21	AAAB43902	Human cancer assoc
721	6	1.7	375	21	AAAB42865	Human ORFX ORF2629	794	6	1.7	413	21	AAAG14116	Arabidopsis thalia
722	6	1.7	375	21	AAAB06449	A human prostate s	795	6	1.7	414	21	AAAG58462	Arabidopsis thalia
723	6	1.7	375	22	AAAB88570	Human hydriophob	796	6	1.7	414	21	AAAB28095	Human secreted pro
724	6	1.7	376	21	AAAG42044	Arabidopsis thalia	797	6	1.7	415	13	AAAB20104	Human secreted pro
725	6	1.7	378	16	AAAG77035	Cat brain-enriched	798	6	1.7	417	21	AAAG56275	Arabidopsis thalia
726	6	1.7	379	20	AAV73874	Human prostate tum	799	6	1.7	417	22	AAAB01209	Arabidopsis thalia
727	6	1.7	379	20	AAV05495	Pseudomonas oxidat	800	6	1.7	419	21	AAAG45957	Human MCH-R2 prote
728	6	1.7	379	21	AAAB56625	Human prostate can	801	6	1.7	420	19	AAV71291	Arabidopsis thalia
729	6	1.7	380	21	AAAG30834	Arabidopsis thalia	802	6	1.7	420	21	AAV81669	Streptococcus pneu
730	6	1.7	381	16	AAAM01433	Glycerol-3-phospha	803	6	1.7	421	21	AAAB57082	Human prostate can
731	6	1.7	385	21	AAV94701	Murine Hzi protein	804	6	1.7	422	21	AAAB12779	Human SLC-1 protei
732	6	1.7	385	21	AAAG14702	Arabidopsis thalia	805	6	1.7	422	21	AAAB13436	Human MCH1 recepto
733	6	1.7	385	21	AAAG42188	Arabidopsis thalia	806	6	1.7	422	21	AAAB13440	Mutant human MCH1
734	6	1.7	386	21	AAV94699	Murine Hzf protein	807	6	1.7	422	21	AAAB13441	Chlamydia pneumoni
735	6	1.7	388	21	AAAG46432	Arabidopsis thalia	808	6	1.7	422	21	AAV92716	Human MCH-R3 prote
736	6	1.7	389	12	AAAR15428	3-acylation enzyme	809	6	1.7	423	22	AAV97670	Human MCH-R3 prote
737	6	1.7	389	16	AAAR80820	Homology vector 52	810	6	1.7	423	21	AAAG37000	Arabidopsis thalia
738	6	1.7	389	21	AAAG34031	Arabidopsis thalia	811	6	1.7	424	18	AAAB18244	PSK-I serine kina
739	6	1.7	389	21	AAAG45812	Arabidopsis thalia	812	6	1.7	427	21	AAAB15967	E. coli proliferat
740	6	1.7	389	22	AAAB69051	Human R2 protein s	813	6	1.7	431	12	AAAR13731	Human cyclin A. H
741	6	1.7	390	20	AAV25115	Human cardiac trop	814	6	1.7	434	16	AAAR75399	Human double minut

815	6	1.7	434	16	AA75496	Human double min
816	6	1.7	434	21	AA858798	Breast and ovarian
817	6	1.7	435	18	AA26131	KOD-1 alpha-amyl
818	6	1.7	436	19	AA878919	Human haemochrom
819	6	1.7	437	21	AA59187	Arabidopsis thalia
820	6	1.7	438	18	AA15092	Human protective p
821	6	1.7	438	20	AA153535	Chlamydia pneumoni
822	6	1.7	438	21	AA30472	Arabidopsis thalia
823	6	1.7	438	21	AA32695	Arabidopsis thalia
824	6	1.7	439	16	AA35742	Arabidopsis thalia
825	6	1.7	439	16	AA85880	WD-40 domain-contg
826	6	1.7	440	18	AA11882	Human p62 polypept
827	6	1.7	440	21	AA35956	Zea mays protein f
828	6	1.7	441	21	AA34115	Arabidopsis thalia
829	6	1.7	441	21	AA34499	Arabidopsis thalia
830	6	1.7	441	21	AA359461	Arabidopsis thalia
831	6	1.7	442	15	AA60536	Mouse osteoblast-s
832	6	1.7	442	16	AA77259	Exo-cellulohydrol
833	6	1.7	443	10	AA93191	Peptide with gluta
834	6	1.7	444	19	AA14940	Corn p-hydroxyphen
835	6	1.7	444	21	AA58933	Breast and ovarian
836	6	1.7	447	21	AA32148	Arabidopsis thalia
837	6	1.7	448	19	AA62351	Human LIR-pm2 pro
838	6	1.7	448	19	AA53463	Human gp49 HML8 po
839	6	1.7	448	21	AA64176	Leukocyte immunogl
840	6	1.7	449	21	AA71285	Streptococcus pneu
841	6	1.7	449	19	AA52286	Gallus domesticus
842	6	1.7	449	21	AA64184	Leukocyte immunogl
843	6	1.7	449	21	AA30230	Arabidopsis thalia
844	6	1.7	450	17	AA89427	Human lymphocyte s
845	6	1.7	451	19	AA38426	Human multiple mye
846	6	1.7	452	18	AA15091	Human precursor pr
847	6	1.7	455	11	AA64576	Polypeptide recogn
848	6	1.7	455	21	AA659433	Arabidopsis thalia
849	6	1.7	456	21	AA63751	Arabidopsis thalia
850	6	1.7	456	21	AA51748	Arabidopsis thalia
851	6	1.7	459	16	AA86729	Aromatic dihydrodi
852	6	1.7	459	21	AA30832	Arabidopsis thalia
853	6	1.7	462	21	AA30471	Arabidopsis thalia
854	6	1.7	462	21	AA31247	Arabidopsis thalia
855	6	1.7	463	20	AA34697	Chlamydia pneumoni
856	6	1.7	463	21	AA635955	Zea mays protein f
857	6	1.7	466	19	AA64357	Aspergillus terreu
858	6	1.7	467	20	AA43169	Consensus phytase-
859	6	1.7	467	20	AA43170	Consensus phytase-
860	6	1.7	467	20	AA43171	A. fumigatus phyta
861	6	1.7	467	20	AA43172	Consensus phytase-
862	6	1.7	467	20	AA19906	Ascomycete consens
863	6	1.7	467	20	AA69380	Fungal phytase pro
864	6	1.7	467	20	AA69381	Fungal consensus p
865	6	1.7	467	20	AA69382	Fungal consensus p
866	6	1.7	467	20	AA69383	Fungal consensus p
867	6	1.7	467	20	AA69384	Fungal consensus p
868	6	1.7	467	20	AA69385	Fungal consensus p
869	6	1.7	467	21	AA620515	Consensus phytase
870	6	1.7	467	21	AA620524	Consensus phytase
871	6	1.7	467	21	AA620526	Consensus phytase
872	6	1.7	467	21	AA620527	Consensus phytase
873	6	1.7	467	21	AA620528	Aspergillus fumiga
874	6	1.7	467	21	AA620529	Consensus phytase
875	6	1.7	467	21	AA620530	Consensus phytase
876	6	1.7	467	21	AA620531	Consensus phytase
877	6	1.7	467	21	AA620532	Consensus phytase
878	6	1.7	467	21	AA620533	Consensus phytase
879	6	1.7	467	21	AA620534	Consensus phytase
880	6	1.7	467	21	AA69558	Phytase-1, a conse
881	6	1.7	467	21	AA695566	Phytase-10, a conse
882	6	1.7	467	21	AA695568	Mutant phytase-1,
883	6	1.7	467	21	AA695569	Mutant phytase-10,
884	6	1.7	467	21	AA695570	Mutant Aspergillus
885	6	1.7	467	21	AA695571	Phytase-12, a deriv
886	6	1.7	467	21	AA695572	ConspHy12, a deriv
887	6	1.7	467	21	AA695574	Mutant Aspergillus

Macaca mulatta rha  
Arabidopsis thalia  
Human G-protein co  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Fcr-IV protein seq  
Rice apoptosis ind  
Enterococcus faeca  
Arabidopsis thalia  
Human 7-transmembr  
Wheat apoptosis in  
Amino acid sequenc  
Corn apoptosis ind  
Rice apoptosis ind  
Human GABAR2 extr  
Synchocystis sp.  
Human lung tumour  
Human developmenta  
Arabidopsis thalia  
Brevibacterium fla  
Chlamydia pneumoni  
Arabidopsis thalia  
Hydrophobic domain  
Human secreted pept  
Human signal pept  
Human MDM2. Homo  
Human MDM2. Homo  
Human MDM2. Invol  
Murine double minu  
Human MDM2. Homo  
Human MDM2 protein  
Human MDM2 protein  
Human MDM2. Homo  
Amino acid sequenc  
Amino acid sequenc  
Human MDM2. Homo  
MDM2 oncprotein.  
Human MDM2 protein.  
Lung cancer associ  
Mouse TANGO 176.  
Arabidopsis thalia  
Human serine/threo  
p70 (beta) S6 Kina  
Human serine/threo  
Human CR1 protein.  
Human src-family k  
Arabidopsis thalia  
T reesei celllobion  
T reesei celllobion  
T reesei celllobion  
Rat mitogenic regu  
Candida tropicalis  
Arabidopsis thalia  
Human GAB5 gene p  
Human Activin rece  
Human ALK-1. Homo  
TAR-3 polypeptide.  
Human activin rece  
Human hARK-1 clone  
Arabidopsis thalia  
Arabidopsis thalia  
Xylanase encoded b  
Aldehyde dehydrogen  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
H. pylori HPS068 p  
H. pylori HPC068 p  
Lactase-LC3. Pol  
Arabidopsis thalia  
Sequence encoded b

961 6 1.7 513 16 AAR77261 T. longibrachiatum  
 962 6 1.7 513 17 AAM02022 Trichoderma celllo  
 963 6 1.7 513 20 AAM94685 Human Del-1 protei  
 964 6 1.7 514 19 AAM57419 Regulatory sequenc  
 965 6 1.7 514 22 AAB79364 Corynebacterium gl  
 966 6 1.7 514 22 AAB49757 Corynebacterium gl  
 967 6 1.7 514 22 AAB49760 Corynebacterium gl  
 968 6 1.7 515 21 AAB56707 Human prostate can  
 969 6 1.7 516 21 AAG39099 Arabidopsis thalia  
 970 6 1.7 520 18 AAM30918 Lily calcium/calmo  
 971 6 1.7 520 21 AAG44022 Arabidopsis thalia  
 972 6 1.7 520 21 AAG54582 Zea mays protein f  
 973 6 1.7 520 22 AAB49336 Zea mays protein f  
 974 6 1.7 521 9 AAP80930 Murine MSP protei  
 975 6 1.7 522 21 AAG44021 Sequence of human  
 976 6 1.7 523 18 AAM22776 Arabidopsis thalia  
 977 6 1.7 523 21 AAG05573 Human septin-2 pro  
 978 6 1.7 523 21 AAG19552 Arabidopsis thalia  
 979 6 1.7 523 22 AAG19552 Human cytoskeleton  
 980 6 1.7 524 21 AAB58954 Alpha-isopropylmal  
 981 6 1.7 524 21 AAY90916 Breast and ovarian  
 982 6 1.7 525 21 AAG54581 Zea mays protein f  
 983 6 1.7 529 17 AAR75847 Arabidopsis adenyl  
 984 6 1.7 531 16 AAR79582 Mouse Cct-zeta sub  
 985 6 1.7 531 21 AAG11571 Arabidopsis thalia  
 986 6 1.7 532 19 AAM49700 Human flavin-conta  
 987 6 1.7 534 20 AAY26966 Wheat Mlo funga  
 988 6 1.7 534 20 AAY26967 Wheat Mlo funga  
 989 6 1.7 534 20 AAY34799 Chlamydia pneumon  
 990 6 1.7 534 21 AAB01805 Wheat Mlo homology  
 991 6 1.7 534 21 AAY67414 Arabidopsis aldehy  
 992 6 1.7 534 22 AAB51250 Amino acid sequenc  
 993 6 1.7 534 22 AAB31251 Human PRO1246 prot  
 994 6 1.7 536 21 AAB33449 Human PRO1246 prot  
 995 6 1.7 536 21 AAB24424 Human PRO1246 prot  
 996 6 1.7 536 21 AAY93377 Human PRO1246 (UNG  
 997 6 1.7 536 22 AAB6126 protein of the inv  
 998 6 1.7 536 22 AAB53097 Human angiogenesis  
 999 6 1.7 537 21 AAG14498 Arabidopsis thalia  
 1000 6 1.7 537 21 AAG54380 Zea mays protein f

## ALIGNMENTS

## RESULT 1

AAM64539 standard; Protein; 350 AA.  
 ID AAM64539

AC AAM64539;

XX 21-OCT-1998 (first entry)

XX Human epidermoid carcinoma cell line KB clone HP10034 protein.

XX Transmembrane domain; human; nutrition; cytokine; cell proliferation;

KM differentiation; immune system; stimulator; suppressor; regulator;

KW haematopoiesis; activin; inhibitor; chemotactic; chemokinetic; receptor;

KM haemostatic; thrombolytic; ligand; anti-inflammatory; tumour.

XX Homo sapiens.

OS WO9821328-A2.

XX 22-MAY-1998.

XX 07-NOV-1997; 97MO-JP04056.

XX 13-NOV-1996; 96JP-0301429.

XX (PROT-) PROTEGENE INC.

PA (SAGA) SAGAMI CHEM RES CENTRE.

XX

PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;  
 XX WPI: 1998-297932/26.  
 DR N-PSDB; AAV49560, AAV49561.  
 XX Human protein having transmembrane domain - useful for, e.g.  
 PT research and nutrition  
 PS Claim 1; Page 98-99; 205pp; English.  
 XX AAM64534-W64558 represent human proteins containing a transmembrane  
 CC domain. These proteins can be used for, e.g. research and nutrition, and  
 CC may have cytokine and cell proliferation/differentiation, immune  
 CC stimulating/suppressing, haematopoiesis regulating, tissue growth,  
 CC activin/inhibin, chemotactic/chemokinetic, haemostatic and thrombolytic,  
 CC receptor/ligand, anti-inflammatory or tumour inhibition activity.  
 XX Sequence 350 AA:

Query Match 100.0%; Score 350; DB 19; Length 350;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 350; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAVFVLLALVAGVILGNEFSILKSPGVSFRNGMNPICGERIPVPAALSGFSVEDLSW 60  
 Db 1 MAVFVLLALVAGVILGNEFSILKSPGVSFRNGMNPICGERIPVPAALSGFSVEDLSW 60  
 QY 61 PGLANGNLFRRPATYVWVWVKVKNKALPFGSVISTYPLENAPPSLDVANSJHSFSEB 120  
 Db 61 PGLANGNLFRRPATYVWVWVKVKNKALPFGSVISTYPLENAPPSLDVANSJHSFSEB 120  
 QY 61 PGLANGNLFRRPATYVWVWVKVKNKALPFGSVISTYPLENAPPSLDVANSJHSFSEB 120  
 Db 61 PGLANGNLFRRPATYVWVWVKVKNKALPFGSVISTYPLENAPPSLDVANSJHSFSEB 120  
 QY 121 TPVVIQLAPEERVYWGKANSVEDLSVTLROLRNRLEFOENSYSSLPINSLSRNNVD 180  
 Db 121 TPVVIQLAPEERVYWGKANSVEDLSVTLROLRNRLEFOENSYSSLPINSLSRNNVD 180  
 QY 181 LRFSELOVLRHDSLSLSRHKHLAKDHPULSYELAGIDELGRYGEDEGFRDASKIL 240  
 Db 181 LRFSELOVLRHDSLSLSRHKHLAKDHPULSYELAGIDELGRYGEDEGFRDASKIL 240  
 QY 241 VDALQKFAEDMYSLYGNAVVELYVKSFTSLIRKRTILAEAKQANPASPYNLAYKYN 300  
 Db 241 VDALQKFAEDMYSLYGNAVVELYVKSFTSLIRKRTILAEAKQANPASPYNLAYKYN 300  
 QY 301 FEYSVFNWVLMIMIALAVITTSYINIMMDPGYDSITTYRRTQKIRMD 350  
 Db 301 FEYSVFNWVLMIMIALAVITTSYINIMMDPGYDSITTYRRTQKIRMD 350

## RESULT 2

AAB90553 standard; Protein; 350 AA.  
 ID AAB90553

AC AAB90553;

XX 01-JUN-2001 (first entry)

XX Human secreted protein, SEQ ID NO: 91.

KM Human; secreted protein; immunomodulatory; antisclerotic;

KW dermatological; antiinflammatory; anti-HIV; cytostatic; cardiant;

KM vascular; anti-angiogenic; ophthalmologically; neutroprotectant;

KW neurotropic; anticonvulsant; antialzheimers; antiparkinsonian;

KM antimicrobial; vulnerary; vaccine; gene therapy; cancer;

XX protein coordinate data; infection.

XX Homo sapiens.

OS WO200121658-A1.

XX 29-MAR-2001.

XX 22-SEP-2000; 2000WO-US26013.

XX 24-SEP-1999; 990S-0155709.  
 PR (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX  
 PI Ni J, Baker KP, Birse CE, Ebner R, Fiscella M, Komatsoulis GA;  
 PI Lafleur DW, Moore PA, Olsen HS, Rosen CA, Ruben SA, Soppet DR;  
 PI Young PE, Wei P, Florence KA;  
 DR WPI: 2001-235311/24.  
 DR N-PSDB; AAF97893.  
 XX  
 PT Nucleic acids encoding 32 human secreted polypeptides, useful for  
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's  
 PT disease and diabetic retinopathy -  
 PS  
 XX Claim 11: Page 776-777; 890pp; English.  
 XX  
 CC The present sequence is one of 32 novel human secreted polypeptides. The  
 CC nucleic acid molecules and polypeptides may be used in the prevention,  
 CC diagnosis and treatment of diseases such as immune disorders (e.g.  
 CC multiple sclerosis, systemic lupus erythematosus and human  
 CC immuno-deficiency virus (HIV) infections), hyperproliferative disorders  
 CC (e.g. cancers and Gaucher's disease), cardiovascular diseases  
 CC (e.g. Schmitz syndrome, Chaga's cardiomyopathy and coronary  
 CC arteriosclerosis), angiogenic disorders (e.g. corneal graft  
 CC neovascularisation and diabetic retinopathy), neurological disorders  
 CC (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),  
 CC infectious diseases and/or for promoting wound healing, regeneration  
 CC and/or chemotaxis. The nucleic acid molecules may be used to produce the  
 CC secreted polypeptides. They may also be used as DNA probes in diagnostic  
 CC assays to detect and quantitate the presence of similar nucleic acid  
 CC sequences in samples. The polypeptides may be used as antigens in the  
 CC production of antibodies and in assays to identify modulators of  
 CC their expression and activity.  
 CC  
 XX  
 SQ Sequence 350 AA:  
 Query Match 100.0%; Score 350; DB 22; Length 350;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 350; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MAVFVLLALVAGVILGNEFSILKSPGSVYFRNGNMPICGERIPDVAALSMGFSVKEDLSW 60  
 DB 1 MAVFVLLALVAGVILGNEFSILKSPGSVYFRNGNMPICGERIPDVAALSMGFSVKEDLSW 60  
 QY 61 PELANGNLFHRPRATVWVWKVGNKIALPPGVSITSPLENAPFSLDSVANSISHSFSE 120  
 DB 61 PELANGNLFHRPRATVWVWKVGNKIALPPGVSITSPLENAPFSLDSVANSISHSFSE 120  
 QY 121 TPVVLQALPSEERYVMGKANSVFEDLSVTLROLNRRLFOENSIVLSLPLNSLRNNEVD 180  
 DB 121 TPVVLQALPSEERYVMGKANSVFEDLSVTLROLNRRLFOENSIVLSLPLNSLRNNEVD 180  
 QY 181 LFTSELQVLDHDSILSRKHLAKDHPOLYSLSLAGDELGRKYGDSERQKDSKTL 240  
 DB 181 LFTSELQVLDHDSILSRKHLAKDHPOLYSLSLAGDELGRKYGDSERQKDSKTL 240  
 QY 241 VDALQKPADWYSLYGNAVVELVYVKSFDTSILRRTIRLEBAKQKNPAPNLAYKYN 300  
 DB 241 VDALQKPADWYSLYGNAVVELVYVKSFDTSILRRTIRLEBAKQKNPAPNLAYKYN 300  
 QY 301 FEYSVFENMVLMTIALALAVITTSYNTNMNDPGYSITIRMTNOKIRMD 350  
 DB 301 FEYSVFENMVLMTIALALAVITTSYNTNMNDPGYSITIRMTNOKIRMD 350  
 QY 301 FEYSVFENMVLMTIALALAVITTSYNTNMNDPGYSITIRMTNOKIRMD 350  
 DB 301 FEYSVFENMVLMTIALALAVITTSYNTNMNDPGYSITIRMTNOKIRMD 350

XX 01-JUN-2001 (first entry)  
 DT  
 XX  
 DE Human secreted protein, SEQ ID NO: 127.  
 XX  
 KW Human: secreted protein; immunomodulatory; antisclerotic;  
 KW dermatological; anti-inflammatory; anti-HIV; cytostatic; cardiant;  
 KW vascular; anti-angiogenic; ophthalmological; neuroprotectant;  
 KW neurotropic; anticonvulsant; antialzheimers; antiparkinsonian;  
 KW antimicrobial; vulnerary; vaccine; gene therapy; cancer;  
 KW protein coordinate data; infection.  
 XX  
 OS Homo sapiens.  
 OS  
 PN WO200121658-A1.  
 XX  
 PD 29-MAR-2001.  
 XX  
 PF 22-SEP-2000; 2000WO-US26013.  
 XX  
 PR 24-SEP-1999; 990S-0155709.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Ni J, Baker KP, Birse CE, Ebner R, Fiscella M, Komatsoulis GA;  
 PI Lafleur DW, Moore PA, Olsen HS, Rosen CA, Ruben SA, Soppet DR;  
 PI Young PE, Wei P, Florence KA;  
 DR WPI: 2001-235311/24.  
 DR N-PSDB; AAF97929.  
 XX  
 PT Nucleic acids encoding 32 human secreted polypeptides, useful for  
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's  
 PT disease and diabetic retinopathy -  
 PS  
 XX Claim 11: Page 809-810; 890pp; English.  
 XX  
 CC The present sequence is one of 32 novel human secreted polypeptides. The  
 CC nucleic acid molecules and polypeptides may be used in the prevention,  
 CC diagnosis and treatment of diseases such as immune disorders (e.g.  
 CC multiple sclerosis, systemic lupus erythematosus and human  
 CC immuno-deficiency virus (HIV) infections), hyperproliferative disorders  
 CC (e.g. cancers and Gaucher's disease), cardiovascular diseases  
 CC (e.g. Schmitz syndrome, Chaga's cardiomyopathy and coronary  
 CC arteriosclerosis), angiogenic disorders (e.g. corneal graft  
 CC neovascularisation and diabetic retinopathy), neurological disorders  
 CC (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),  
 CC infectious diseases and/or for promoting wound healing, regeneration  
 CC and/or chemotaxis. The nucleic acid molecules may be used to produce the  
 CC secreted polypeptides. They may also be used as DNA probes in diagnostic  
 CC assays to detect and quantitate the presence of similar nucleic acid  
 CC sequences in samples. The polypeptides may be used as antigens in the  
 CC production of antibodies and in assays to identify modulators of  
 CC their expression and activity.  
 CC  
 XX  
 SQ Sequence 350 AA:  
 Query Match 98.9%; Score 346; DB 22; Length 350;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 5 VVILALVAGVILGNEFSILKSPGSVYFRNGNMPICGERIPDVAALSMGFSVKEDLSW 64  
 DB 5 VVILALVAGVILGNEFSILKSPGSVYFRNGNMPICGERIPDVAALSMGFSVKEDLSW 64  
 QY 65 VGNLFHRPRATVWVWKVGNKIALPPGVSITSPLENAPFSLDSVANSISHSFSEPTV 124  
 DB 65 VGNLFHRPRATVWVWKVGNKIALPPGVSITSPLENAPFSLDSVANSISHSFSEPTV 124  
 QY 125 LQALPSEERYVMGKANSVFEDLSVTLROLNRRLFOENSIVLSLPLNSLRNNEVDLFT 184  
 DB 125 LQALPSEERYVMGKANSVFEDLSVTLROLNRRLFOENSIVLSLPLNSLRNNEVDLFT 184

RESULT 3  
 AAB90589  
 ID AAB90589 standard; Protein: 350 AA.  
 XX  
 AC AAB90589;

QY 185 SELQVLDHDSLSLRHKLAKDHPDLYSLAELDEIGKRGYGEDEQFRDASKILVDL 244  
 |||||  
 Db 185 selqvlhdsllsrhklakdhsdpdlyslaeldeigkrygeseqfrdaskilvda1 244  
 QY 245 QKFADDMYSLXGNAVVELYTKSFDTSILKTRITILEKQKKNAPSPYNAIKYNEFS 304  
 |||||  
 Db 245 qkfaddmyslygnavveltykvsfdtsilkrtrllleakqaknpaspylnaykynfeys 304  
 QY 305 VVENAVLWIMIALALAVITTSYNNMDPGYDSITTYRMTNOKIRMD 350  
 |||||  
 Db 305 vvenavlwimialalavitsyynnmdpgydsityrmtncqirmd 350

RESULT 4  
 AAB88347  
 ID AAB88347 standard; Protein: 350 AA.  
 AC AAB88347;  
 XX  
 DT 23-MAY-2001 (first entry)  
 XX  
 DE Human membrane or secretory protein clone PSEC0072.  
 XX  
 KW Human; secretory protein; membrane protein; vaccine; gene therapy;  
 KM rheumatoid arthritis; diabetes.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1067182-A2.  
 XX  
 PD 10-JAN-2001.  
 XX  
 PF 07-JUL-2000; 2000EP-0114090.  
 XX  
 PR 08-JUL-1999; 99JP-0194179.  
 PR 11-JAN-2000; 2000JP-0118775.  
 PR 02-MAY-2000; 2000JP-0183766.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;  
 DR WPI; 2001-093989/11.  
 DR N-PSDB; AAF93774.  
 XX  
 PT Nucleic acids encoding secretory proteins/membrane proteins, useful in  
 PT gene therapy or as candidate target molecules in drug development -  
 XX  
 PS Claim 1; SEQ ID 62; 609pp + CD ROM; English.  
 XX  
 CC This invention relates to nucleic acid sequences AAF93744 - AAF93916  
 CC which encode human secretory or membrane proteins represented by  
 CC AAB88317 - AAB88419. Included in the invention are primers  
 CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the  
 CC cDNA sequences of the invention. The invention also includes methods for  
 CC the production of antibodies directed against the proteins, and cDNA  
 CC sequences, which can be used in vaccines. The polynucleotide sequences  
 CC can be used in gene therapy. The polynucleotide sequences and the  
 CC proteins they encode may be used in the prevention, treatment and  
 CC diagnosis of diseases associated with inappropriate secretory  
 CC protein/membrane protein expression. The nucleic acids and complementary  
 CC sequences may also be used as DNA probes in diagnostic assays  
 CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the  
 CC presence of similar nucleic acid sequences in samples. They may also be  
 CC used to study the expression and function of secretory proteins/membrane  
 CC polypeptides and their role in metabolism. The polypeptides may be used  
 CC as antigens in the production of antibodies against them and in assays to  
 CC identify modulators (agonists and antagonists) of expression and  
 CC activity. The antibodies and antagonists may also be used as therapeutic  
 CC agents to down regulate expression and activity. The antibodies may also  
 CC be used as diagnostic agents for detecting the presence of the  
 CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay

CC (ELISA). Examples of diseases which may be treated include rheumatoid  
 CC arthritis and diabetes.  
 XX  
 SO Sequence 350 AA:

Query Match 73.4%; Score 257; DB 22; Length 350;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-237;  
 Matches 257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAVFVLLAIVAGVIGNESTILKSPGVSVPFRNGNMPFRGERIPVVAALSMGSKYVEDLSW 60  
 |||||  
 Db 1 mavfvllaivagvlgnefstlkspgsvvfrngnmpfrgeripvvaalsmgskysvedls 60  
 QY 61 PGLAVGNLFRRPRATVWVWVGKVKLALPPGSVISTPLENAPVPSLDVSANSHISFSEE 120  
 |||||  
 Db 61 pglavgnlfrpratvwvwvgkvkllalppgsvisvtpelenavpplsldvsanshlsfsee 120  
 QY 121 TPVYVQLAPSEERVYWGKANSVFEDLSYTLRLRLRFLFOENSVLSSLPLNSLRNNVYD 180  
 |||||  
 Db 121 tpvyvqlapseervwvgkansvfedlsytlrlnrlfogensvlsstplnslsrnnv 180  
 QY 181 LLFLSELQVLDHDSLSLRHKLAKDHPDLYSLAELDEIGKRGYGEDEQFRDASKIL 240  
 |||||  
 Db 181 llflselqvlhdsllsrhklakdhsdpdlyslaeldeigkrygeseqfrdaskil 240  
 QY 241 VDAIQKFIADDMYSLY99 257  
 |||||  
 Db 241 vdaiqkfiadmysly99 257

RESULT 5  
 AAY59720  
 ID AAY59720 standard; Protein: 335 AA.  
 XX  
 AC AAY59720;  
 XX  
 DT 18-JAN-2000 (first entry)  
 XX  
 DE Secreted protein 33-77-4-E2-PL1.  
 XX  
 DE  
 XX  
 KW Secreted protein; fingerprint identification technique;  
 KW chromosome mapping; human; hereditary disease; diagnosis; cancer;  
 KW hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;  
 KW autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;  
 KW renal injury; amino aciduria; hypoglycaemia; male rat infertility;  
 KW hypertension.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9940189-A2.  
 XX  
 PD 12-AUG-1999.  
 XX  
 PF 09-FEB-1999; 99WO-IB00282.  
 XX  
 PR 09-FEB-1998; 98US-0074121.  
 PR 13-APR-1998; 98US-0081563.  
 PR 10-AUG-1998; 98US-0096116.  
 PR 04-SEP-1998; 98US-0099273.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Bougueleret L, Duclert A, Dumas Milne Edwards J;  
 DR WPI; 1999-60096/51.  
 DR N-PSDB; AAZ40848.  
 XX  
 PT Extended cDNAs useful for expressing secreted proteins and to obtain  
 PT specific antibodies -  
 XX  
 PS Claim 10; Page 241; 244pp; English.  
 XX



CC This sequence represents a human secreted protein of the invention.  
 CC The extended cDNAs (or genomic DNAs obtainable from them) may be used to  
 CC prepare PCR primers and probes. These are useful for forensic matching or  
 CC positive identification by DNA sequencing. They may also be used in  
 CC alternative fingerprint identification techniques. Antibodies against the  
 CC proteins encoded by the extended cDNAs are useful in identification of  
 CC tissue types or cell species, as well as identifying tissue specific  
 CC soluble proteins. The sequences can be used for chromosome mapping and  
 CC identification of genes associated with hereditary diseases or drug  
 CC response. Signal sequences from the cDNAs can be used in construction of  
 CC secretion vectors. Other sequences derived from the extended cDNAs can be  
 CC used to clone upstream genomic DNA sequences including promoters. This is  
 CC in turn useful for identifying proteins that interact with promoter  
 CC sequences. Some of the proteins may be useful in diagnosing and treating  
 CC several disorders including, but not limited to: cancer, hyperlipidaemia,  
 CC cardiovascular and neurodegenerative disorders, autoimmune diseases, and  
 CC rheumatic diseases, embryogenic disorders, hypertension, renal injury,  
 CC amino acidurias, hypoglycaemia, male rat infertility and myopathies.  
 CC  
 XX Sequence 335 AA:

Query Match 42.9%; Score 150; DB 20; Length 335;  
 Best Local Similarity 100.0%; Pred. No. 2,7e-135;  
 Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 182 LFLSEQLQVLDISSLLSRKHLAKDHPDLYSLELAGDEIGKRYGDSDFDASKILV 241  
 DB 182 LFLSEQLQVLDISSLLSRKHLAKDHPDLYSLELAGDEIGKRYGDSDFDASKILV 241  
 OY 242 DALQKADDMYSLYGNAVVELVTKSPFSLIRKRTTLEKQAKNPASPYMLAKYKF 301  
 DB 242 DALQKADDMYSLYGNAVVELVTKSPFSLIRKRTTLEKQAKNPASPYMLAKYKF 301  
 OY 302 EYSVVENMVLIMIALALAVITTSNTNM 331  
 DB 302 EYSVVENMVLIMIALALAVITTSNTNM 331

## RESULT 6

AAV13118  
 ID AAV13118 standard; Protein: 93 AA.

AC AAV13118;

DT 22-JUN-1999 (first entry)

DE Human secreted protein encoded by 5' EST SEQ ID NO: 132.

XX Human; secreted protein; EST; expressed sequence tag; diagnosis;  
 KW forensic; gene therapy; chromosome mapping; signal peptide;  
 KW upstream regulatory sequence; cytokine activity; cell proliferation;  
 KW differentiation; haematopoiesis regulation; tissue growth regulation;  
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; anti-inflammatory; tumour inhibition.

OS Homo sapiens.

PN WO906552-A2.

PD 11-FEB-1999.

PF 31-JUL-1998; 98WO-IB01236.

PR 01-AUG-1997; 97US-0905223.

PA (GEST ) GENSET.

PI Duclert A, Dumas Milne Edwards J, Lacroix B;

XX WPI: 1999-153782/13.

DR N-PSDB; AAX51918.

PT New isolated brain-derived nucleic acids - used to develop products  
 PT which may have cytokine, immune, regulatory, haematopoiesis  
 PT regulating, anti-inflammatory or tumour inhibition activity  
 XX  
 XX Claim 34: Page 516; 577pp; English.

CC AAX51787 to AAX52019 represent 5' expressed sequence tags (ESTs) for  
 CC human secreted proteins, and encode the proteins given in AAV12987 to  
 CC AAV13219, respectively. The proteins given represent the signal peptide  
 CC and an N-terminal fragment of a secreted protein. The nucleic acid  
 CC sequences can be used for producing secreted human gene products. They  
 CC can also be used to develop products for diagnosis and therapy. The  
 CC proteins obtained may have cytokine activity, cell  
 CC proliferation/differentiation activity, haematopoiesis regulating  
 CC activity, tissue growth regulating activity, reproductive hormone  
 CC regulating activity, chemotactic/chemokinetic activity, haemostatic and  
 CC thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, tumour inhibition activity or other activities. The products  
 CC can be used in forensic, gene therapy and chromosome mapping procedures.  
 CC The sequences can also be used for obtaining corresponding promoter  
 CC sequences. The nucleic acids encoding the signal peptide can be used for  
 CC directing extracellular secretion of a polypeptide or the insertion of a  
 CC polypeptide into a membrane, or importing a polypeptide into a cell.  
 CC  
 XX Sequence 93 AA:

Query Match 21.1%; Score 74; DB 20; Length 93;  
 Best Local Similarity 100.0%; Pred. No. 4.3e-63;  
 Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MAVFVLLIAYAGVIGNEPSILKSGSVFRNGMPIGERIPVVALNSGFSVKEDLSW 60  
 DB 1 MAVFVLLIAYAGVIGNEPSILKSGSVFRNGMPIGERIPVVALNSGFSVKEDLSW 60  
 OY 61 PGLAVGNLFRRPRA 74  
 DB 61 PGLAVGNLFRRPRA 74

## RESULT 7

AAV12371  
 ID AAV12371 standard; Protein: 139 AA.

AC AAV12371;

DT 17-JUN-1999 (first entry)

DE Human 5' EST secreted protein SEQ ID NO:402.

XX Human; secreted protein; EST; expressed sequence tag; diagnosis;  
 KW forensic; gene therapy; chromosome mapping; signal peptide;  
 KW upstream regulatory sequence; cytokine activity; cell proliferation;  
 KW differentiation; haematopoiesis regulation; tissue growth regulation;  
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; anti-inflammatory; tumour inhibition.

OS Homo sapiens.

PN WO906548-A2.

PD 11-FEB-1999.

PF 31-JUL-1998; 98WO-IB01222.

PR 01-AUG-1997; 97US-0905135.

PA (GEST ) GENSET.

PI Duclert A, Dumas Milne Edwards J, Lacroix B;

XX WPI: 1999-153778/13.

DR N-PSDB; AAX41204.

XX New nucleic acids encoding human secreted proteins - obtained from  
PT CDNA libraries prepared from e.g. liver, ovary, brain, prostate,  
PT kidney, lung, umbilical cord, placenta and colon tissue  
XX  
PS Claim 27; Page 724-725; 824pp; English.  
XX  
CC AAX11094 to AAX11347 represent 5' expressed sequence tags (ESTs) for  
CC human secreted proteins, and encode the proteins given in AAX11261 to  
CC AAX12514, respectively. The proteins given represent the signal peptide  
CC and an N-terminal fragment of a secreted protein. The nucleic acid  
CC sequences can be used for producing secreted human gene products. They  
CC can also be used to develop products for diagnosis and therapy. The  
CC proteins obtained may have cytokine activity, cell  
CC proliferation/differentiation activity, haematopoiesis regulating  
CC activity, tissue growth regulating activity, reproductive hormone  
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and  
CC thrombolytic activity, receptor/ligand activity anti-inflammatory  
CC activity, tumour inhibition activity or other activities. The products  
CC can be used in forensic, gene therapy and chromosome mapping procedures.  
CC The sequences can also be used for obtaining corresponding promoter  
CC sequences. The nucleic acids encoding the signal peptide can be used for  
CC directing extracellular secretion of a polypeptide or the insertion of a  
CC polypeptide into a membrane, or importing a polypeptide into a cell.  
XX  
SQ Sequence 139 AA:

Query Match 2.3%, Score 8; DB 20; Length 139;  
Best Local Similarity 100.0%; Pred. No. 8.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 LILFLESLQ 188  
|||||||  
Db 40 ILFLESLQ 47

RESULT 8  
AAV71106  
ID AAV71106 standard; Protein: 153 AA.  
XX  
AC AAV71106;  
XX  
DT 08-SEP-2000 (first entry)  
XX  
DE Human Hyaluronase protein-4 (HYDRL-4).  
XX  
XX Hyaluronase; HYDRL; human; clone 1553276; cytosolic; immunosuppressive;  
XX antiinflammatory; neuroprotective; cerebroprotective; anticonvulsant;  
XX nephroprotective; antibody; agonist; antagonist; diagnosis; treatment;  
XX prevention; cell proliferative disorder; cancer; inflammatory; AIDS;  
XX Acquired Immune Deficiency Syndrome; autoimmune/inflammatory disorder;  
XX neurological disorder; epilepsy; stroke; medullary sponge kidney;  
XX renal disorder; adrenal disorder; adrenoleucodystrophy.  
XX  
XX Homo sapiens.  
XX  
OS  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..21  
FT /label= Signal\_peptide  
FT Protein 22..153  
FT /note= "Mature human hyaluronase-4 protein"  
FT Modified-site 31  
FT /note= "Phosphorylation site"  
XX  
XX WO200028045-A2.  
XX  
XX 18-MAY-2000.  
XX  
XX 12-NOV-1999; 99WO-US27009.  
XX  
XX 12-NOV-1998; 98US-0172256.  
XX  
XX 21-MAY-1999; 99US-0135519.

XX (INCYTE) INCYTE PHARM INC.  
XX  
XX Tang TY, Hallman JL, Yue H, Lai P, Bandman O, Corley NC;  
XX Guegler KJ, Baughn MR, Lu DAM, Azimzal Y, Yang J;  
XX  
XX WPI: 2000-376557/32.  
XX N-PSDB: AAD00679.  
XX  
PT Novel human hyaluronase protein useful for diagnosing, treating and  
PT preventing cell proliferative, autoimmune and inflammatory,  
PT neurological, renal, adrenal and genetic disorders  
XX  
XX Claim 1; Page 80; 106pp; English.  
XX  
XX The present sequence is the human hyaluronase protein-4 (HYDRL-4),  
XX identified in Incyte clone 1553276, derived from the BLADTUT04 library.  
XX It is expressed in urologic tissues and has homology to Collipase from  
XX Myocastor coypus.  
XX HYDRL antibodies, agonists and antagonists are useful for diagnosis,  
XX treatment and prevention of disorders associated with altered expression  
XX or activity of HYDRL. It includes cell proliferative disorder such as  
XX cancer, autoimmune or inflammatory disorders such as inflammation, AIDS,  
XX neurological disorder such as epilepsy, stroke, renal disorder such as  
XX medullary sponge kidney, adrenal disorder such as adrenoleucodystrophy.  
XX The nucleotide sequences are useful for detecting and quantifying gene  
XX expression in tissues.  
XX  
SQ Sequence 153 AA:

Query Match 2.3%, Score 8; DB 21; Length 153;  
Best Local Similarity 100.0%; Pred. No. 9.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LALVAGYL 15  
|||||||  
Db 5 lalvagyl 12

RESULT 9  
AAV13942  
ID AAV13942 standard; Protein: 383 AA.  
XX  
AC AAV13942;  
XX  
DT 14-JUL-1999 (first entry)  
XX  
DE Human transmembrane protein, HP01737.  
XX  
XX  
XX Transmembrane protein; human; cell membrane; proliferation; diagnosis;  
XX cell differentiation; carcinostatic agent; probe; gene therapy;  
XX signal transduction; apoptosis; inhibitor;  
XX phosphatidylinethanolamine N-methyltransferase.  
XX  
XX  
XX Homo sapiens.  
XX  
OS  
XX  
XX WO9918203-A2.  
XX  
XX 15-APR-1999.  
XX  
XX 05-OCT-1998; 98WO-JP04475.  
XX  
XX 08-OCT-1997; 97JP-0276271.  
XX  
XX (PROT-) PROTEGENE INC.  
XX (SAGA ) SAGAMI CHEM RES CENT.  
XX  
XX Kato S, Kobayashi M, Sekine S, Yamaguchi T;  
XX  
XX WPI: 1999-277268/23.  
XX  
XX N-PSDB: AAX36808, AAX36809.

PT Human transmembrane proteins and nucleotide sequences  
 XX  
 PS Claim 1; Page 87-89; 139pp; English.  
 CC This sequence is a human transmembrane protein of the invention.  
 CC All of the proteins exist in the cell membrane, so are considered to be  
 CC proteins controlling the proliferation and differentiation of the cells.  
 CC They may be useful as carciostatic agents or as antigens for preparing  
 CC antibodies against the proteins. The CDNA's can be used as probes for  
 CC gene diagnosis and gene sources for gene therapy, as well as for  
 CC large-scale expression of the proteins. The HP01498 (see AA13339)  
 CC protein may be associated with signal transduction associated with  
 CC apoptosis, and therefore useful in inhibition of apoptosis. The HP01962  
 CC (see AA13943) protein can be used to treat diseases associated with  
 CC phosphatidylethanolamine N-methyltransferase. The proteins are  
 CC identified by the presence of a hydrophobic transmembrane region,  
 CC knowledge of the protein function is not required, as in e.g. methods of  
 CC expression cloning.  
 CC  
 SQ Sequence 383 AA;  
 XX  
 XX  
 Query Match 2.3%; Score 8; DB 20; Length 383;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 181 LFLFSEIQ 188  
 DB 40 ILLFSEIQ 47  
 DE  
 RESULT 10  
 ID AAB36611 standard; Protein: 389 AA.  
 AC AAB36611;  
 DT 09-MAR-2001 (first entry)  
 XX  
 DE Human FLEXHT-33 protein sequence SEQ ID NO:33.  
 KW Human; FLEXHT; full-length molecules expressed in human tissue;  
 KW diagnosis; gene expression; genetic linkage; genetic variability;  
 KW antianaemic; anticonvulsant; antiarteriosclerotic; immunomodulatory;  
 KW cytosolic; antiasthmatic; antiinflammatory; hepatotropic; antidiabetic;  
 KW anti-gout; antihypertoid; neuroprotective; antiarthritic; osteopathic;  
 KW antipsoriatic; antirheumatic; antitumor; gene therapy; anaemia; gout;  
 KW epilepsy; arteriosclerosis; atherosclerosis; developmental disorder;  
 KW cancer; immunological disorder; asthma; bronchitis; cirrhosis;  
 KW Crohn's disease; diabetes mellitus; Grave's disease; multiple sclerosis;  
 KW osteoarthritis; pancreatitis; rheumatoid arthritis; psoriasis;  
 KW ulcerative colitis.  
 KW  
 XX  
 OS Homo sapiens.  
 XX  
 PN WC200070047-A2.  
 XX  
 PD 23-NOV-2000.  
 XX  
 PF 12-MAY-2000; 2000MO-US13299.  
 XX  
 PR 14-MAY-1999; 99US-0311894.  
 PR 14-MAY-1999; 99US-0311937.  
 PR 14-MAY-1999; 99US-0311940.  
 XX  
 PA (INCYTE GENOMICS INC.  
 XX  
 PI Yue H, Tang YH, Lal P, Reddy R, Batria S, Baughn MK, Yang J;  
 PI Azimzai Y, Lu DAM, Au-Young J, Shih LL;  
 DR WPI, 2001-016234/02.  
 DR N-PSDB; AAC88102.  
 XX

PT Human FLEXHT protein and DNA sequences, useful for treating  
 PT Immunological disorders, developmental disorders, and cancers -  
 XX  
 PS Claim 1; Page 121-122; 168pp; English.  
 CC AAC88070 to AAC88124 encode the 55 FLEXHT (full-length molecules  
 CC expressed in human tissues) proteins given in AAB36579 to AAB36633. The  
 CC present invention describes an isolated polypeptide (A) comprising an  
 CC amino acid sequence selected from one of 55 amino acid sequences 42-876  
 CC residues in length, corresponding to FLEXHT-1 to FLEXHT-55, a 90 %  
 CC identical sequence, and a biologically active or immunogenic fragment of  
 CC the sequence. The FLEXHT proteins can have antianaemic, anticonvulsant,  
 CC antiarteriosclerotic, immunomodulatory, cytosolic, antiasthmatic,  
 CC antiinflammatory, hepatotropic, antidiabetic, anti-gout, antihypertoid,  
 CC neuroprotective, antiarthritic, osteopathic, antipsoriatic, antitumor  
 CC and antirheumatic activities, and can be used in gene therapy. The  
 CC polynucleotide sequences can be used to express the protein sequences.  
 CC Pharmaceutical compositions comprising FLEXHT can be used to treat  
 CC diseases or conditions associated with altered expression of functional  
 CC FLEXHT. The proteins and polynucleotides can be used to diagnose and  
 CC treat disorders including anaemia, epilepsy, arteriosclerosis,  
 CC atherosclerosis, developmental disorders, cancers, and immunological  
 CC disorders such as asthma, bronchitis, cirrhosis, Crohn's disease,  
 CC diabetes mellitus, gout, Grave's disease, multiple sclerosis,  
 CC osteoarthritis, pancreatitis, psoriasis, rheumatoid arthritis, and  
 CC ulcerative colitis.  
 CC  
 SQ Sequence 389 AA;  
 XX  
 XX  
 Query Match 2.3%; Score 8; DB 22; Length 389;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 181 LFLFSEIQ 188  
 DB 40 ILLFSEIQ 47  
 DE  
 RESULT 11  
 ID AAG56388 standard; Protein: 38 AA.  
 AC AAG56388;  
 DT 18-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 72475.  
 KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 KW  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN EP1033405-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 25-FEB-2000; 2000EP-0301439.  
 XX  
 PR 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 23-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 990S-0130510.  
PR 23-APR-1999; 990S-0130891.  
PR 28-APR-1999; 990S-0131449.  
PR 30-APR-1999; 990S-0132048.  
PR 30-APR-1999; 990S-0132407.  
PR 04-MAY-1999; 990S-0132484.  
PR 05-MAY-1999; 990S-0132485.  
PR 06-MAY-1999; 990S-0132486.  
PR 07-MAY-1999; 990S-0132487.  
PR 11-MAY-1999; 990S-0134256.  
PR 14-MAY-1999; 990S-0134218.  
PR 14-MAY-1999; 990S-0134219.  
PR 14-MAY-1999; 990S-0134221.  
PR 14-MAY-1999; 990S-0134370.  
PR 18-MAY-1999; 990S-0134768.  
PR 19-MAY-1999; 990S-0134941.  
PR 20-MAY-1999; 990S-0135124.  
PR 21-MAY-1999; 990S-0135353.  
PR 24-MAY-1999; 990S-0135629.  
PR 25-MAY-1999; 990S-0136021.  
PR 27-MAY-1999; 990S-0136352.  
PR 28-MAY-1999; 990S-0136782.  
PR 01-JUN-1999; 990S-0137222.  
PR 03-JUN-1999; 990S-0137528.  
PR 04-JUN-1999; 990S-0137502.  
PR 07-JUN-1999; 990S-0137724.  
PR 08-JUN-1999; 990S-0138094.  
PR 10-JUN-1999; 990S-0138540.  
PR 10-JUN-1999; 990S-0138647.  
PR 14-JUN-1999; 990S-0139119.  
PR 16-JUN-1999; 990S-0139452.  
PR 16-JUN-1999; 990S-0139453.  
PR 17-JUN-1999; 990S-0139452.  
PR 18-JUN-1999; 990S-0139454.  
PR 18-JUN-1999; 990S-0139455.  
PR 18-JUN-1999; 990S-0139456.  
PR 18-JUN-1999; 990S-0139457.  
PR 18-JUN-1999; 990S-0139458.  
PR 18-JUN-1999; 990S-0139459.  
PR 18-JUN-1999; 990S-0139460.  
PR 18-JUN-1999; 990S-0139461.  
PR 18-JUN-1999; 990S-0139462.  
PR 18-JUN-1999; 990S-0139463.  
PR 18-JUN-1999; 990S-0139750.  
PR 18-JUN-1999; 990S-0139763.  
PR 21-JUN-1999; 990S-0139817.  
PR 22-JUN-1999; 990S-0139899.  
PR 23-JUN-1999; 990S-0140353.  
PR 23-JUN-1999; 990S-0140354.  
PR 24-JUN-1999; 990S-0140695.  
PR 28-JUN-1999; 990S-0140823.  
PR 29-JUN-1999; 990S-0140991.  
PR 30-JUN-1999; 990S-0141287.  
PR 01-JUL-1999; 990S-0141842.  
PR 01-JUL-1999; 990S-0142154.  
PR 02-JUL-1999; 990S-0142055.  
PR 06-JUL-1999; 990S-0142390.  
PR 08-JUL-1999; 990S-0142803.  
PR 09-JUL-1999; 990S-0142920.  
PR 12-JUL-1999; 990S-0142977.  
PR 13-JUL-1999; 990S-0143542.  
PR 14-JUL-1999; 990S-0143624.  
PR 15-JUL-1999; 990S-0144005.  
PR 16-JUL-1999; 990S-0144086.  
PR 16-JUL-1999; 990S-0144086.  
PR 19-JUL-1999; 990S-0144325.  
PR 19-JUL-1999; 990S-0144331.  
PR 19-JUL-1999; 990S-0144332.  
PR 19-JUL-1999; 990S-0144333.  
PR 19-JUL-1999; 990S-0144334.  
PR 19-JUL-1999; 990S-0144335.  
PR 20-JUL-1999; 990S-0144352.

PR 20-JUL-1999; 990S-0144632.  
PR 20-JUL-1999; 990S-0144684.  
PR 21-JUL-1999; 990S-0144814.  
PR 21-JUL-1999; 990S-0145086.  
PR 21-JUL-1999; 990S-0145088.  
PR 22-JUL-1999; 990S-0145085.  
PR 22-JUL-1999; 990S-0145087.  
PR 22-JUL-1999; 990S-0145089.  
PR 22-JUL-1999; 990S-0145192.  
PR 23-JUL-1999; 990S-0145145.  
PR 23-JUL-1999; 990S-0145224.  
PR 26-JUL-1999; 990S-0145276.  
PR 27-JUL-1999; 990S-0145913.  
PR 27-JUL-1999; 990S-0145918.  
PR 27-JUL-1999; 990S-0145919.  
PR 28-JUL-1999; 990S-0145951.  
PR 02-AUG-1999; 990S-0146386.  
PR 02-AUG-1999; 990S-0146388.  
PR 02-AUG-1999; 990S-0146389.  
PR 03-AUG-1999; 990S-0147038.  
PR 04-AUG-1999; 990S-0147204.  
PR 04-AUG-1999; 990S-0147302.  
PR 05-AUG-1999; 990S-0147192.  
PR 05-AUG-1999; 990S-0147260.  
PR 06-AUG-1999; 990S-0147303.  
PR 06-AUG-1999; 990S-0147416.  
PR 09-AUG-1999; 990S-0147493.  
PR 09-AUG-1999; 990S-0147935.  
PR 10-AUG-1999; 990S-0148171.  
PR 11-AUG-1999; 990S-0148341.  
PR 12-AUG-1999; 990S-0148341.  
PR 13-AUG-1999; 990S-0148565.  
PR 13-AUG-1999; 990S-0148565.  
PR 16-AUG-1999; 990S-0149368.  
PR 17-AUG-1999; 990S-0149175.  
PR 18-AUG-1999; 990S-0149426.  
PR 20-AUG-1999; 990S-0149722.  
PR 20-AUG-1999; 990S-0149723.  
PR 20-AUG-1999; 990S-0149929.  
PR 23-AUG-1999; 990S-0149902.  
PR 23-AUG-1999; 990S-0149930.  
PR 25-AUG-1999; 990S-0150566.  
PR 26-AUG-1999; 990S-0150884.  
PR 27-AUG-1999; 990S-0151065.  
PR 27-AUG-1999; 990S-0151066.  
PR 30-AUG-1999; 990S-0151080.  
PR 31-AUG-1999; 990S-0151303.  
PR 01-SEP-1999; 990S-0151938.  
PR 01-SEP-1999; 990S-0151930.  
PR 07-SEP-1999; 990S-0152363.  
PR 10-SEP-1999; 990S-0153070.  
PR 13-SEP-1999; 990S-0153758.  
PR 15-SEP-1999; 990S-0154018.  
PR 16-SEP-1999; 990S-0154039.  
PR 20-SEP-1999; 990S-0154779.  
PR 22-SEP-1999; 990S-0155139.  
PR 23-SEP-1999; 990S-0155486.  
PR 24-SEP-1999; 990S-0155659.  
PR 28-SEP-1999; 990S-0156458.  
PR 29-SEP-1999; 990S-0156596.  
PR 04-OCT-1999; 990S-0157117.  
PR 05-OCT-1999; 990S-0157753.  
PR 06-OCT-1999; 990S-0157865.  
PR 07-OCT-1999; 990S-0158029.  
PR 08-OCT-1999; 990S-0158232.  
PR 12-OCT-1999; 990S-0158369.  
PR 13-OCT-1999; 990S-0159293.  
PR 13-OCT-1999; 990S-0159284.  
PR 13-OCT-1999; 990S-0159295.  
PR 14-OCT-1999; 990S-0159329.  
PR 14-OCT-1999; 990S-0159330.  
PR 14-OCT-1999; 990S-0159331.

PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160815.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160988.  
PR 22-OCT-1999; 99US-0160988.  
PR 23-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 2.0%; Score 7; DB 21; Length 38;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 163 SVLSSLP 169  
DB 27 swisslp 33

## RESULT 12

ID AAB38132 standard; Protein: 42 AA.

AC AAB38132;

DT 30-JAN-2001 (first entry)

DE Human secreted protein sequence encoded by gene 14 SEQ ID NO:71.

XX Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;  
XX antirheumatic; antiproliferative; cytostatic; cardiact; vasotrophic;  
XX cerebroprotective; neurotropic; antibacterial; virucide;  
XX fungicide; ophthalmological; gene therapy; autoimmune disease; infection;  
XX hyperproliferative disorder; cardiovascular disorder; angiogenesis;  
XX cerebrovascular disorder; nervous system disorder; ocular disorder;  
XX wound healing; skin aging; food additive; preservative.

OS Homo sapiens.

PN WO200058468-A2.

PD 05-OCT-2000.

PF 22-MAR-2000; 2000WO-US07526.

PR 26-MAR-1999; 99US-0126600.

PR 22-DEC-1999; 99US-0171550.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM, Komatsoulis G;

DR MPI: 2000-611713/58.

DR N-PSDB; AAC69412.

XX Nucleic acids encoding human secreted proteins, used to prevent, treat,  
PT ameliorate, or diagnose conditions such as autoimmune disorders, skin  
XX disorders, and cancer -

PS Claim 11; Page 347; 374pp; English.

XX The polynucleotide sequences given in AAC69399 to AAC69445 encode the  
CC human secreted proteins given in AAB38119 to AAB38165. AAB38166 to  
CC AAB38201 represent human secreted polypeptide sequences and proteins  
CC homologous to them, which are given in the exemplification of the present  
CC invention. Human secreted proteins have activities based on the tissues  
CC and cells the genes are expressed in. Example of activities include:  
CC immunosuppressive; antirheumatic; antiproliferative;  
CC cytostatic; cardiact; vasotrophic; cerebroprotective; neurotropic;  
CC neuroprotective; antibacterial; virucide; fungicide; and  
CC ophthalmological. The polynucleotides and polypeptides can be used to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used  
CC in diagnosing a pathological condition or susceptibility to a  
CC pathological condition. Disorders which are diagnosed or treated include  
CC autoimmune diseases, hyperproliferative disorders, cardiovascular  
CC disorders, cerebrovascular disorders, angiogenesis, nervous system  
CC disorders, infections caused by bacteria, viruses and fungi and ocular  
CC disorders. The polypeptides can also be used to aid wound healing and  
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to  
CC maintain organs before transplantation, for supporting cell culture of  
CC primary tissues, to regenerate tissues and in chemotaxis. The  
CC polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities. AAC69399 to AAC69398 and  
CC AAB38118 represent sequences used in the exemplification of the present  
CC invention.

XX Sequence 42 AA;

Query Match 2.0%; Score 7; DB 21; Length 42;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 181 LFLFSEL 187

DB 4 1111111 10

## RESULT 13

ID AAC20221 standard; Protein: 61 AA.

AC AAC20221;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 22326.

XX Protein identification; signal transduction pathway; metabolic pathway;  
XX hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145102.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149308.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153708.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.

PR 14-OCT-1999; 99US-0158637.  
 PR 14-OCT-1999; 99US-0158638.  
 PR 18-OCT-1999; 99US-0158584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161358.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161921.  
 PR 28-OCT-1999; 99US-0161933.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 2.0%; Score 7; DB 21; Length 61;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 163 SVLSLIP 169  
 |||||  
 DB 49 svlselp 55

RESULT 14  
 AAR07630  
 ID AAR07630 standard; peptide: 66 AA.  
 XX  
 AC AAR07630;  
 XX  
 DT 10-APR-1991 (first entry)  
 XX  
 DE N-terminal of Fraction I polypeptide from spider venom.  
 XX  
 KM Calcium channel blocker; angina; hypertension; Raynaud's disease.  
 XX  
 OS Agelenopsis aperta.  
 XX  
 PN EP395357-A.  
 XX  
 PD 31-OCT-1990.  
 XX  
 PF 24-APR-1990; 90EP-0304397.  
 XX  
 PR 28-APR-1989; 89US-0346181.  
 XX  
 PA (PF12 ) PFIZER INC.  
 PA (NATU-) NATURAL PRODUCTS SCIENCES INC.  
 XX  
 PI Saccamano NA, Volkman RA;  
 DR WPI; 1990-329348/44.  
 XX  
 PT New poly-amine(s) and polypeptide(s) from spider venom - are  
 PT excitatory amino acid neuro-transmitter antagonists and calcium  
 PT channel blockers.  
 XX  
 PS Claim 2; Page 43; 54pp; English.  
 XX  
 CC The sequence represents the N-terminal of a peptide present in the  
 CC crude venom of the Agelenopsis aperta spider. It elutes off a C-18  
 CC Vydac (RTM) 22mm x 250mm, 300A pore size, 10u particle size column  
 CC using a flow rate of 15 ml/min and a linear gradient of 5-20% B,  
 CC 95-80% A (0-30 mins), then 20-70% B, 80-30% A (30-55 mins), where  
 CC A = acetonitrile and B = 0.1% aq. TFA. The polypeptide elutes in

CC fraction I at 40 mins. It also elutes off a similar column (C-4)  
 CC at 22 mins if a flow rate of 10 ml/min and a non-linear gradient of  
 CC 20-30% B, 80-70% A (0-30 mins), then 30-50% B, 70-45 mins)  
 CC is used. The molecular weight of the protein is 4158 (FAB MS).  
 CC The polypeptide blocks calcium channels in nerves and muscles  
 CC and can be used to treat e.g. angina, hypertension, cardiomyopathies,  
 CC etc. It can also be used to control invertebrate pests.  
 CC See also AAR07628-33.  
 XX  
 SQ Sequence 66 AA;

Query Match 2.0%; Score 7; DB 11; Length 66;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 87 ALPGSV 93  
 |||||  
 DB 3 alppsv 9

RESULT 15  
 AAR38420  
 ID AAR38420 standard; polypeptide: 66 AA.  
 XX  
 AC AAR38420;  
 XX  
 DT 21-DEC-1993 (first entry)  
 XX  
 DE A. aperta venom fraction H2 calcium channel blocker.  
 XX  
 KM Spider venom; treatment; angina; hypertension; seizure; stroke;  
 KM cardiomyopathies; supraventricular arrhythmia; oesophageal achalasia;  
 KM premature labour; Raynaud's disease; invertebrate pest control;  
 KM cerebral ischaemia; neuronal degenerative disorders; epilepsy;  
 KM Alzheimer's disease; psychotherapeutics; study; cell physiology.  
 XX  
 OS Agelenopsis aperta.  
 XX  
 FH Key Location/Qualifiers  
 FT Disulfide-bond 50 /note= "cys-S-S-cys-pro-ser-NH2"  
 XX  
 PN US5227397-A.  
 XX  
 PD 13-JUL-1993.  
 XX  
 PF 17-JUL-1990; 90US-0554311.  
 XX  
 PR 28-APR-1989; 89US-0346181.  
 PR 14-MAR-1990; 90US-0491066.  
 PR 17-JUL-1990; 90US-0554311.  
 XX  
 PA (PF12 ) PFIZER INC.  
 PA Saccamano NA, Volkman RA;  
 PI WPI; 1993-235157/29.  
 DR  
 XX  
 PT Indole 3-acetyl derivs. of poly(amine(s)) - useful as calcium channel  
 PT blockers and neuro-transmitter antagonists  
 XX  
 PS Disclosure; Page 4; 28pp; English.  
 XX  
 CC The sequence is that of a polypeptide present in fraction H2 of the  
 CC venom of Agelenopsis aperta. It blocks calcium channels in cells  
 CC of both mammals and invertebrates, partic. those affecting neuronal  
 CC and muscle cells. It may be used in the treatment of angina,  
 CC hypertension, cardiomyopathies, supraventricular arrhythmia,  
 CC oesophageal achalasia, premature labour, and Raynaud's disease.  
 CC It may also be of use in the study of cell physiology and in the  
 CC control of invertebrate pests. It may be produced synthetically.

SQ Sequence 66 AA:

Query Match 2.0%; Score 7; DB 14; Length 66;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 87 ALPGSV 93  
 |||||  
 Db 3 alpgsv 9

RESULT 16

AAR8582

ID AAR8582 standard; Peptide: 66 AA.

AC AAR8582;

DT 19-JUL-1996 (first entry)

DE Spider venom calcium channel blocker (7793 Da) N-terminal sequence.

KW Spider venom; calcium channel; antagonist; blocker;

KW cardiovascular disease; pesticide.

OS Agelenopsis aperta.

PN EP696578-A2.

PD 14-FEB-1996.

PF 24-APR-1990; 90EP-0105540.

PR 28-APR-1989; 89US-0346181.

PA (NPSP-) NPS PHARM INC.

PA (Pfiz ) PFIZER INC.

PI Saccomano NA, Volkman RA;

PI WPI, 1996-098575/11.

PT Polyamine and polypeptide calcium channel blockers from A. aperta

PT venom - for the treatment of cardiovascular diseases and controlling

PT insect pests

PS Claim 2; Page 49; 57pp; English.

CC The N-terminal sequence (AAR8582) was detd. of a protein obtd. from  
 CC the venom of the A. aperta spider. The protein was isolated by  
 CC fractionation of the venom on a C-18 Ydac (RPM) column. It has a  
 CC predicted mol.wt. of 7793 (ion-spray MS). The protein, and other  
 CC venom proteins (see also AAR8580-81 and AAR8583-85), block calcium  
 CC channels in a variety of cells, such as those of the nervous and  
 CC muscular system of invertebrates and vertebrates, including mammals.  
 CC They are useful for treating cardiovascular diseases, for studying  
 CC cell physiology, and for the control of invertebrate pests.

SQ Sequence 66 AA:

Query Match 2.0%; Score 7; DB 17; Length 66;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 87 ALPGSV 93  
 |||||  
 Db 3 alpgsv 9

RESULT 17

AAB66709  
 ID AAB66709 standard; protein: 83 AA.

XX AAB66709;

DT 09-APR-2001 (first entry)

DE C.glutamicum phosphoenolpyruvate protein #3.

KW Phosphoenolpyruvate; sugar phosphotransferase system; PTS.

OS Corynebacterium glutamicum.

PN WO200102583-A2.

PD 11-JAN-2001.

PF 27-JUN-2000; 2000WO-1B00973.

PR 01-JUL-1999; 99US-0142691.

PR 23-AUG-1999; 99US-0150310.

PR 03-SEP-1999; 99DE-1042095.

PR 03-SEP-1999; 99DE-1042097.

PA (BADI ) BASF AG.

PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;

PI WPI, 2001-080989/09.

PT Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate;

PT sugar phosphotransferase system proteins or their portions, useful for

PT typing or identifying C. glutamicum or related bacteria, and as markers

PT for transformation

PS Claim 8; Page 106; 144pp; English.

CC The present invention relates to Corynebacterium glutamicum

CC phosphoenolpyruvate; sugar phosphotransferase system (PTS) proteins.

CC The PTS nucleic acids and proteins are useful in the

CC identification of microorganisms which can be used to produce fine

CC chemicals, for modulating fine chemical production in C. glutamicum or

CC related bacteria, the typing or identification of C. glutamicum or

CC related bacteria, as reference points for mapping C. glutamicum genome,

CC and as markers for transformation.

SQ Sequence 83 AA;

Query Match 2.0%; Score 7; DB 22; Length 83;  
 Best Local Similarity 100.0%; Pred. No. 50;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 122 PVIQLA 128  
 |||||  
 Db 66 pviqla 72

RESULT 18

AAY76593  
 ID AAY76593 standard; Protein: 119 AA.

AC AAY76593;

DT 10-APR-2000 (first entry)

DE Human ovarian tumor EST fragment encoded protein 89.

KW Expressed sequence tag; EST; human; ovarian tumor; anticancer;

KW gene therapy; treatment.

OS Homo sapiens.

XX DE19817557-A1.





CC (AAW19604) (MG075). They are homologues of 16 and 116 kDa proteins  
CC (see also AAW19601-02) obtd. from Mycoplasma pneumoniae. A genomic  
CC DNA sequence of M. genitalium contains contiguous open reading  
CC frames that code for the 2 polypeptides. Mycoplasma 16 or 116 kDa  
CC proteins, or immunogenic fragments that include a T or B cell  
CC epitope, can be used in vaccines for prevention and treatment of  
CC Mycoplasma infections, partic. in humans. They can also be used  
CC diagnostically to detect Mycoplasma, or to raise antibodies useful  
CC in immunoassays.  
XX  
SQ Sequence 137 AA;  
  
Query Match 2.0%; Score 7; DB 18; Length 137;  
Best Local Similarity 100.0%; Pred. No. 79;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 181 LRLSEL 187  
Db 14 LRLSEL 20  
  
RESULT 21  
AAG18462  
ID AAG18462 standard; Protein; 153 AA.  
XX  
AC AAG18462;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 19880.  
XX  
KW Protein identification: signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 990US-0121825.  
PR 05-MAR-1999; 990US-0123180.  
PR 09-MAR-1999; 990US-0123548.  
PR 23-MAR-1999; 990US-0125788.  
PR 25-MAR-1999; 990US-0126284.  
PR 29-MAR-1999; 990US-0126785.  
PR 01-APR-1999; 990US-0127462.  
PR 06-APR-1999; 990US-0128234.  
PR 08-APR-1999; 990US-0128714.  
PR 16-APR-1999; 990US-0129845.  
PR 19-APR-1999; 990US-0130077.  
PR 21-APR-1999; 990US-0130449.  
PR 23-APR-1999; 990US-0130510.  
PR 23-APR-1999; 990US-0130891.  
PR 28-APR-1999; 990US-0131449.  
PR 30-APR-1999; 990US-0132048.  
PR 30-APR-1999; 990US-0132407.  
PR 04-MAY-1999; 990US-0132484.  
PR 05-MAY-1999; 990US-0132485.  
PR 06-MAY-1999; 990US-0132486.  
PR 06-MAY-1999; 990US-0132487.  
PR 07-MAY-1999; 990US-0132863.  
PR 11-MAY-1999; 990US-0134256.  
PR 14-MAY-1999; 990US-0134218.  
PR 14-MAY-1999; 990US-0134219.  
PR 14-MAY-1999; 990US-0134221.  
PR 14-MAY-1999; 990US-0134370.  
PR 18-MAY-1999; 990US-0134768.  
PR 19-MAY-1999; 990US-0134941.

PR 20-MAY-1999; 990US-0135124.  
PR 21-MAY-1999; 990US-0135353.  
PR 24-MAY-1999; 990US-0135629.  
PR 25-MAY-1999; 990US-0136021.  
PR 27-MAY-1999; 990US-0136392.  
PR 28-MAY-1999; 990US-0136782.  
PR 01-JUN-1999; 990US-0137222.  
PR 03-JUN-1999; 990US-0137528.  
PR 04-JUN-1999; 990US-0137702.  
PR 07-JUN-1999; 990US-0137724.  
PR 08-JUN-1999; 990US-0138094.  
PR 10-JUN-1999; 990US-0138540.  
PR 10-JUN-1999; 990US-0138847.  
PR 14-JUN-1999; 990US-0139119.  
PR 16-JUN-1999; 990US-0139452.  
PR 16-JUN-1999; 990US-0139453.  
PR 17-JUN-1999; 990US-0139492.  
PR 18-JUN-1999; 990US-0139454.  
PR 18-JUN-1999; 990US-0139455.  
PR 18-JUN-1999; 990US-0139456.  
PR 18-JUN-1999; 990US-0139457.  
PR 18-JUN-1999; 990US-0139458.  
PR 18-JUN-1999; 990US-0139459.  
PR 18-JUN-1999; 990US-0139460.  
PR 18-JUN-1999; 990US-0139461.  
PR 18-JUN-1999; 990US-0139462.  
PR 18-JUN-1999; 990US-0139463.  
PR 18-JUN-1999; 990US-0139750.  
PR 18-JUN-1999; 990US-0139763.  
PR 21-JUN-1999; 990US-0139817.  
PR 22-JUN-1999; 990US-0139899.  
PR 23-JUN-1999; 990US-0140353.  
PR 23-JUN-1999; 990US-0140354.  
PR 24-JUN-1999; 990US-0140695.  
PR 28-JUN-1999; 990US-0140823.  
PR 29-JUN-1999; 990US-0140981.  
PR 30-JUN-1999; 990US-0141287.  
PR 01-JUL-1999; 990US-0141842.  
PR 01-JUL-1999; 990US-0142154.  
PR 02-JUL-1999; 990US-0142055.  
PR 06-JUL-1999; 990US-0142390.  
PR 08-JUL-1999; 990US-0142803.  
PR 09-JUL-1999; 990US-0142920.  
PR 12-JUL-1999; 990US-0142977.  
PR 13-JUL-1999; 990US-0143542.  
PR 14-JUL-1999; 990US-0143624.  
PR 15-JUL-1999; 990US-0144005.  
PR 16-JUL-1999; 990US-0144085.  
PR 16-JUL-1999; 990US-0144086.  
PR 19-JUL-1999; 990US-0144325.  
PR 19-JUL-1999; 990US-0144331.  
PR 19-JUL-1999; 990US-0144332.  
PR 19-JUL-1999; 990US-0144333.  
PR 19-JUL-1999; 990US-0144334.  
PR 19-JUL-1999; 990US-0144335.  
PR 20-JUL-1999; 990US-0144352.  
PR 20-JUL-1999; 990US-0144632.  
PR 20-JUL-1999; 990US-0144884.  
PR 21-JUL-1999; 990US-0144814.  
PR 21-JUL-1999; 990US-0145086.  
PR 21-JUL-1999; 990US-0145088.  
PR 22-JUL-1999; 990US-0145087.  
PR 22-JUL-1999; 990US-0145089.  
PR 22-JUL-1999; 990US-0145089.  
PR 22-JUL-1999; 990US-0145192.  
PR 23-JUL-1999; 990US-0145145.  
PR 23-JUL-1999; 990US-0145218.  
PR 23-JUL-1999; 990US-0145224.  
PR 26-JUL-1999; 990US-0145276.  
PR 27-JUL-1999; 990US-0145913.  
PR 27-JUL-1999; 990US-0145918.  
PR 27-JUL-1999; 990US-0145919.  
PR 28-JUL-1999; 990US-0145951.

PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156536.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.

PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 2.0%; Score 7; DB 21; Length 153;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 4 FVLLAL 10  
Db 19 fvvllal 25  
  
RESULT 22  
ACGI8461  
ID ACGI8461 standard; Protein; 160 AA.  
AC ACGI8461;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 19879.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
XX  
PE 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0123788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142305.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145142.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151085.  
PR 27-AUG-1999; 99US-0151086.  
PR 27-AUG-1999; 99US-0151087.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match

2.0%; Score 7; DB 21; Length 160;

Best Local Similarity 100.0%; Pred. No. 91;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 FVLLAL 10  
Db 26 fvllal 32

## RESULT 23

AA089581  
ID AA089581 standard; Protein; 175 AA.

XX  
AC AA089581;

DT 25-APR-1996 (first entry)

XX Partial bg28 insert B-G subregion fragment of fowl MHC.

DE Domesticated fowl; chicken; turkey; pheasants; B-G antigen; MHC;

KW major histocompatibility complex; haplotyping; Marek's disease;

XX restriction fragment length polymorphism; bg28 insert; probe.

OS Synthetic.

XX US5451670-A.

XX 19-SEP-1995.

XX 30-JUN-1987; 87US-0068176.

XX 22-APR-1991; 91US-0688326.

XX 30-JUN-1987; 87US-0068176.

XX 09-DEC-1987; 87US-0130529.

XX 23-JUN-1988; 88US-0210405.

XX 28-SEP-1989; 89US-0413301.

XX 27-SEP-1990; 90US-0588922.

XX 07-APR-1992; 92US-0865662.

XX (CITY ) CITY OF HOPE.

XX MILLER MM;

XX WPI: 1995-336319/43.

XX N-PSDB: AA089944.

XX New DNA encoding the B-G antigen of fowl MHC - and derived probes,

XX useful for haplotyping and determining genotype at specific loci

XX Claim 5; Columns 13-14; 61pp; English.

XX The DNA sequences AA089944-55 contain a polymorphic portion of the

XX coding region of domestic fowl's (i.e. chicken, turkey, pheasant)

XX MHC B-G subregion (AA089944 encodes the B-G subregion fragment

XX AA089581). The sequences are used as probes to detect restriction

XX fragment length polymorphism patterns typical of various B-G

XX alleles, e.g. resistance to Marek's (and other) disease, general

XX fitness and productivity, all are related to MHC haplotype. This

XX haplotyping method has the advantage of not requiring antisera.

XX Sequence 175 AA;

ID AA093200 standard; Protein; 180 AA.  
XX  
XX AA093200;  
AC  
XX  
XX 08-NOV-2000 (first entry)  
DT  
XX  
XX Human connective tissue growth factor protein fragment.  
DE  
XX  
XX Dermatological; antiarthritic; antiarteriosclerotic; antidiabetic;  
KW nephrotropic; ophthalmological; hypotensive; cardiac; tranquilizer;  
KW vulnery; antiinflammatory; human; connective tissue growth factor;  
KW CTGF; extracellular matrix synthesis; collagen synthesis; antibody;  
KW myofibroblast differentiation; antisense; fibroproliferative disease;  
KW fibrosis; trauma; cancer; inflammation; diabetes; ketoid.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200035936-A1.  
XX  
XX 22-JUN-2000.  
PD  
XX  
XX 14-DEC-1999; 99WO-US29652.  
PF  
XX  
XX 14-DEC-1998; 98US-0112240.  
PR  
XX  
XX 14-DEC-1998; 98US-0112241.  
PR  
XX  
XX (UYMI-) UNIV MIAMI.  
PA  
XX  
XX Grolendorst GR;  
PI  
XX  
XX WPI: 2000-431565/37.  
DR  
XX  
XX N-PSDB: AAA11279.  
DR  
XX  
XX Fragment of connective tissue growth factor, useful for treating  
PT fibroproliferative diseases or disorders, including kidney fibrosis,  
PT scleroderma, arthritis, hypertrophic scarring, atherosclerosis, diabetic  
PT neuropathy and retinopathy -  
PT  
XX  
XX Disclosure; Fig 4; 74pp; English.  
PS  
XX  
XX This sequence represents a fragment of the human connective tissue  
XX growth factor (CTGF) polypeptide having the ability to induce  
XX extracellular matrix synthesis, collagen synthesis and/or myofibroblast  
XX differentiation. The protein covers the amino acids encoded by exons 2  
XX and 3 of the full length cDNA sequence (AAA11278). The invention relates  
XX to fragments of CTGF, especially those encoded by exons 2 and/or 3,  
XX which contain the inductive activity. The protein and/or fragments can be  
XX used to raise antibodies and the coding sequence can be used to generate  
XX antisense oligonucleotides. The antibody or antisense sequence against  
XX the CTGF sequence can be used in a method to treat a CTGF-associated  
XX disease or disorder such as a fibroproliferative disease or disorder,  
XX especially selected from kidney fibrosis, scleroderma, pulmonary  
XX fibrosis, liver fibrosis, arthritis, hypertrophic scarring,  
XX atherosclerosis, diabetic nephropathy and retinopathy, hypertension,  
XX kidney disorders, angiogenesis-related disorders, skin fibrotic  
XX disorders, and cardiovascular disorders. The disease or disorder can  
XX also be selected from acute or repetitive traumas (including surgery or  
XX radiation therapy, and fibrosis of organs), diseases caused by vascular  
XX endothelial cell proliferation or migration (including cancers),  
XX inflammatory bowel disease, Crohn's disease, joint inflammation,  
XX interstitial disease, dermatological diseases, diabetes, and ketoids.  
XX  
XX Sequence 180 AA;

Query Match 2.0%; Score 7; DB 21; Length 180;

Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 FVLLAL 10  
Db 13 fvllal 19

RESULT 24  
AA093200

RESULT 25  
ID AAB40410 standard; Protein: 188 AA.  
AC AAB40410;  
DT 08-FEB-2001 (first entry)  
XX  
XX  
DE Human ORFX ORF174 polypeptide sequence SBQ ID NO:348.  
XX  
XX  
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnery; antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;  
KW anticonvulsant; osteopathic; antileukemic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antineoplastic; antileukemic;  
KW antiviral; antibacterial; antifungal; antineoplastic; antileukemic;  
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive.  
XX  
XX  
OS Homo sapiens.  
XX  
XX  
PN WO200058473-A2.  
XX  
PD 05-OCT-2000.  
XX  
PE 31-MAR-2000; 2000WO-US08621.  
XX  
PF 31-MAR-1999; 99US-0127607.  
PR 02-APR-1999; 99US-0127636.  
PR 05-APR-1999; 99US-0127728.  
PR 30-MAR-2000; 2000US-0540763.  
XX  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
XX  
PI Shinkets RA, Leach M;  
XX  
XX  
DR WPI: 2000-602362/57.  
DR N-PSDB; AAC74619.  
XX  
XX  
PT Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX  
XX  
PS Claim 11; Page 610-611; 5507pp; English.  
XX  
XX  
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
CC antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;  
CC osteopathic; anticonvulsant; antileukemic; immunosuppressant;  
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antineoplastic; antibacterial; antiviral; antifungal; antineoplastic;  
CC antileukemic; antianemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy.  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX  
XX  
SQ Sequence 188 AA:  
XX  
XX  
Query Match 2.0%; Score 7; DB 21; Length 188;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Caps 0;  
XX  
XX  
OY 315 IALALAV 321  
DB 44 IALALAV 50  
XX  
XX  
RESULT 26  
AAC15696  
ID AAC15696 standard; Protein: 203 AA.  
XX  
AC AAC15696;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
XX  
DE Arabidopsis thaliana protein fragment SBQ ID NO: 16051.  
XX  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
XX  
OS Arabidopsis thaliana.  
XX  
XX  
PN EP103405-A2.  
XX  
XX  
PD 06-SEP-2000.  
XX  
PE 25-FEB-2000; 2000EP-0301439.  
XX  
PF 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123160.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0128845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134376.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137122.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139753.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 28-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 03-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149358.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151086.  
 PR 30-AUG-1999; 99US-0151080.  
 PR 31-AUG-1999; 99US-0151303.  
 PR 01-SEP-1999; 99US-0151438.  
 PR 07-SEP-1999; 99US-0151930.  
 PR 10-SEP-1999; 99US-0152363.  
 PR 13-SEP-1999; 99US-0153070.  
 PR 15-SEP-1999; 99US-0153758.  
 PR 16-SEP-1999; 99US-0154018.  
 PR 20-SEP-1999; 99US-0154039.  
 PR 22-SEP-1999; 99US-0154779.  
 PR 23-SEP-1999; 99US-0155139.  
 PR 24-SEP-1999; 99US-0155486.  
 PR 26-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 2.0%; Score 7; DB 21; Length 203;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 SLEPNSL 173  
Db 70 slpnsl 76  
RESULT 27  
AAG53950  
ID AAG53950 standard: Protein: 203 AA.  
AC AAG53950;  
XX 18-OCT-2000 (first entry)  
DT Arabidopsis thaliana protein fragment SEQ ID NO: 68734.  
DE  
XX  
XX Protein identification: signal transduction pathway: metabolic pathway:  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
OS  
XX  
XX EPI033405-A2.  
PN  
XX  
XX 06-SEP-2000.  
PD  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
PE  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-013629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.



PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0158029.  
 PR 07-OCT-1999; 99US-0158232.  
 PR 08-OCT-1999; 99US-0158368.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0158293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 14-OCT-1999; 99US-0159584.  
 PR 18-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161922.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 2.0%; Score 7; DB 21; Length 203;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 SLPINSL 173  
 |||||  
 Db 70 slpInsl 76

RESULT 28

AA061570  
 ID AA061570 standard; Protein: 203 AA.  
 AC AA061570;  
 XX  
 XX 18-OCT-2000 (first entry)  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 79873.  
 XX  
 KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX Arabidopsis thaliana.  
 OS  
 XX EPI033405-A2.  
 PD  
 XX 06-SEP-2000.  
 XX  
 XX 25-FEB-2000; 2000EP-0301439.  
 PF  
 XX 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123160.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 04-MAY-1999; 99US-0132407.  
 PR 05-MAY-1999; 99US-0132484.  
 PR 06-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 07-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 03-JUN-1999; 99US-0137222.  
 PR 04-JUN-1999; 99US-0137528.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 17-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 22-JUN-1999; 99US-0139817.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142927.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 19-JUL-1999; 99US-0144336.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145088.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.

PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## Query Match

Best Local Similarity 2.08; Score 7; DB 21; Length 203;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 167 SLPLNSL 173  
Db 70 slplns 76

## RESULT 29

AAV51680  
ID AAV51680 standard; Protein; 218 AA.

XX AAV51680;

AC AAV51680;  
XX  
DT 02-JUN-2000 (first entry)  
XX

DE Murine clodin 5 protein.  
XX  
KW Clodin 5; murine; tight junction-constituting membrane protein;  
KW medicine.  
XX  
OS Mus sp.  
PN JP2000032984-A.  
XX  
XX 02-FEB-2000.  
XX  
XX 26-JUN-1998; 98JP-0179847.  
XX  
XX 15-MAY-1998; 98JP-0133215.  
XX  
XX (EISA ) EISAI CO LTD.  
XX  
XX WPI; 2000-285512/25.  
XX  
XX N-PSDB; AA289152.  
XX  
XX Tight junction-constituting membrane protein clodin family - useful in  
PT the medical field  
XX  
PS Claim 4; Page 16-17; 22pp; Japanese.  
XX  
XX This invention describes novel murine nucleic acid sequences encoding the  
CC clodin family of tight junction (TJ)-constituting membrane protein. The  
CC membrane protein can be used in medical field. This sequence represents  
CC the clodin 5 protein described in the method of the invention.  
XX  
SQ Sequence 218 AA:  
  
Query Match 2.08; Score 7; DB 21; Length 218;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 6 VILALVA 12  
DB 88 VILALVA 94  
  
RESULT 30  
AAG09413  
ID AAG09413 standard; Protein: 289 AA.  
XX  
AC AAG09413;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 7338.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
XX  
XX 09-MAR-1999; 99US-0123180.  
XX  
XX 23-MAR-1999; 99US-0123548.  
XX  
XX 25-MAR-1999; 99US-0125788.  
XX  
XX 29-MAR-1999; 99US-0126264.  
XX  
XX 01-APR-1999; 99US-0126785.  
XX  
XX 06-APR-1999; 99US-0127462.  
XX  
XX 08-APR-1999; 99US-0128234.  
XX  
XX 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136382.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0138119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0138458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0138817.  
PR 22-JUN-1999; 99US-0138899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142927.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147453.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149358.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151086.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 13-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156586.  
PR 04-OCT-1999; 99US-0157717.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0159284.  
PR 13-OCT-1999; 99US-0159295.  
XX

PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 2.08; Score 7; DB 21; Length 289;  
Best Local Similarity 100.08; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 193 ISSLSR 199  
Db 260 ISSLSR 266  
  
RESULT 31  
AAV38843  
ID AAV38843 standard; Protein: 307 AA.  
XX  
AC AAV38843;  
XX  
DT 08-OCT-1999 (first entry)  
XX  
DE Neisseria meningitidis antigen encoded by a partial ORF26.  
XX  
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
treatment; Neisseria infection; meningitis; septicemia; gonorrhea.  
XX  
OS Neisseria meningitidis.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 98..99 \*197 unspecified amino acids are present  
FT /note= between these residues"  
FT  
PN WO9924578-A2.  
XX  
PD 20-MAY-1999.  
XX  
PF 09-OCT-1998; 98WO-IB01665.  
XX  
PR 01-SEP-1998; 98GB-0019016.  
PR 06-NOV-1997; 97GB-0023516.  
PR 14-NOV-1997; 97GB-0024190.  
PR 18-NOV-1997; 97GB-0024386.  
PR 27-NOV-1997; 97GB-0025158.  
PR 10-DEC-1997; 97GB-0026147.  
PR 14-JAN-1998; 98GB-0000759.  
XX  
PA (CHIR-) CHIRON SPA.  
XX  
PI Grandi G, Massignani V, Piza M, Rappuoli R, Scarlato V;  
XX

DR WPI: 1999-327407/27.  
DR P-PSDB: AAY3843.  
XX  
PT Proteins from *Neisseria meningitidis* and *N. gonorrhoeae* useful for  
PT diagnosis, treatment and prevention of infection  
XX  
PS Claim 4; Page 394; 524pp; English.  
XX  
CC Amino acid sequences AAY38499-Y38944 represent *Neisseria meningitidis*  
CC and *N. gonorrhoeae* antigenic proteins. They are encoded by open  
CC reading frames (ORFs) AA21972-21358. The antigenic proteins,  
CC their fragments, their nucleic acids and antibodies are used for  
CC diagnosis, prevention (as vaccines) or treatment of *Neisseria*  
CC infections, such as meningitis, septicaemia and gonorrhoea. Both  
CC organisms are closely related. Fragments of the nucleic acids  
CC are useful as hybridisation probes and antisense reagents.  
CC  
SQ Sequence 307 AA:

Query Match 2.0% Score 7; DB 20; Length 307;  
Best Local Similarity 100.0% Pred No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 316 ALALAVI 322  
|||||  
DB 19 alalavi 25

RESULT 32  
AAC09412  
ID AAC09412 standard; Protein: 310 AA.

AC AAC09412;

DT 17-OCT-2000 (First entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 7337.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

PN EPI033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-012180.

PR 03-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 08-APR-1999; 99US-0128234.

PR 16-APR-1999; 99US-0128714.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132485.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139839.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142803.  
PR 12-JUL-1999; 99US-0142920.  
PR 13-JUL-1999; 99US-0142977.  
PR 14-JUL-1999; 99US-0143542.  
PR 15-JUL-1999; 99US-0143624.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144633.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 03-AUG-1999; 99US-0146389.  
 PR 04-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 05-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 06-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 09-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 16-AUG-1999; 99US-0148684.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 23-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151920.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159687.  
 PR 14-OCT-1999; 99US-0159688.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.

PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 2.0%; Score 7; DB 21; Length 310;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 ISSLSR 199  
 DB 261 ISSLSR 287

RESULT 33  
 AAM12694

ID AAM12694 standard: Protein; 347 AA.

XX AAM12694:

DT 31-MAY-1997 (first entry)

DE Connective tissue growth factor.

KW Connective tissue growth factor; CTGF; bone; cartilage; vulnary;

XX wound healing; osteoporosis; osteoarthritis; osteochondrytis.

OS Mus sp.

PN M09638166-A1.

PD 05-DEC-1996.

PF 31-MAY-1996; 96WO-US08210.

PR 31-MAY-1996; 96WO-US08140.

PA (GRO1) GROTEENDORST G R.

PI Groteendorst GR;

DR WPI: 1997-042658/04.

DR N-PSDB: AAT59618.

PT Connective Tissue Growth Factor composition - for inducing bone,  
 tissue and cartilage formation and wound healing

XX Disclosure: Fig 1C1-3; 60pp; English.

PS Connective tissue growth factor (CTGF) (AAM12694) is a cysteine-rich  
 CC mitogenic protein which is selectively induced in fibroblasts after  
 CC activation with transforming growth factor beta (TGF-beta).  
 CC Recombinant CTGF can be produced in prokaryotic or eukaryotic host  
 CC cells utilizing an isolated CTGF gene (AAT59618). Compens.  
 CC comprising CTGF, pref. in combination with TGF-beta, are used to  
 CC induce bonding formation, e.g. to treat osteoporosis,  
 CC osteoarthritis and osteochondrytis, to induce tissue and cartilage  
 CC formation, and to induce wound healing. It can also be used in  
 CC culture systems e.g. to expand stem cells or chondrocytes prior to  
 CC re-implantation. CTGF is more stable to protease degradation than  
 CC other growth factors used as prior art wound healing agents.

SO Sequence 347 AA;

Query Match 2.0%; Score 7; DB 18; Length 347;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 FVVLAL 10  
 |||||  
 Db 13 fvvllal 19

RESULT 34  
 AAR79964  
 ID AAR79964 standard; Protein: 349 AA.

XX AAR79964;  
 XX  
 DT 12-JUN-1996 (first entry)

XX Connective tissue growth factor.

XX Connective tissue growth factor; CTGF; wound healing; vulnerary;  
 KW cell proliferation; cancer; fibrosis; atherosclerosis; diagnosis  
 KM therapy; mitogen.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Modified-site 28 /label= N-glycosylation\_site

FT Modified-site 225 /label= N-glycosylation\_site

FT US5408040-A.

XX 18-APR-1995.

XX 30-AUG-1991; 91US-0752427.

XX 30-AUG-1991; 91US-0752427.

XX 14-DEC-1993; 93US-0167628.

XX (UYSF-) UNIV SOUTH FLORIDA.

XX Bradham DM, Grotendorst GR;

XX WPI: 1995-161147/21.

XX N-PSDB; AAT04226.

PT New connective tissue growth factor - used to develop prods. for  
 PT wound healing and for diagnosis and therapy of cell proliferative  
 PT disorders.

XX Claim 1; Column 19-20; 12pp; English.

XX Novel human connective tissue growth factor (CTGF) (AAR79964)

XX is related immunologically and biologically to platelet-derived

XX growth factor (PDGF), but is the product of a distinct gene.

XX CTGF is mitogenic and also a chemotactic agent for cells. It is

XX produced by endothelial and fibroblastic cells, and probably acts

XX as a growth factor in wound healing. Recombinant CTGF can be outd.

XX by expression of cDNA clone DB60R32 (AAT04226) in transformed host

XX cells. It is used to accelerate wound healing, and to raise

XX antibodies useful in detecting disorders associated with overgrowth

XX of cells, such as cancer, fibrotic diseases and atherosclerosis.

XX Sequence 349 AA;

Query Match 2.0%; Score 7; DB 16; Length 349;

Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 FVVLAL 10  
 |||||

Db 13 fvvllal 19

RESULT 35  
 AAM09089  
 ID AAM09089 standard; Protein: 349 AA.

XX AAM09089;

XX 26-APR-1997 (first entry)

XX Human connective tissue growth factor.

XX Connective tissue growth factor; CTGF; mitogen; cell proliferation;  
 KW wound healing; cancer; tumour; fibrosis; glaucoma; atherosclerosis;  
 KM scleroderma; arthritis, cirrhosis, scar; diagnosis; therapy.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Modified-site 28 /label= Glycosylation

FT /note= "potential N-glycosylation site"

FT Modified-site 225 /label= Glycosylation

FT /note= "potential N-glycosylation site"

FT W09638172-A1.

XX 05-DEC-1996.

XX 31-MAY-1996; 96WO-US08140.

XX 31-MAY-1996; 96WO-US08140.

XX (UYSF-) UNIV SOUTH FLORIDA.

XX Bradham DM, Grotendorst GR;

XX WPI: 1997-042659/04.

XX N-PSDB; AAT45360.

XX N-PSDB; AAT58534.

PT Connective tissue growth factor coding sequence and protein - used  
 PT in the treatment of proliferative disorders and to accelerate wound  
 PT healing

XX Claim 19; Page 50-52; 76pp; English.

XX Novel human connective tissue growth factor (CTGF) (AAM09089) is a

XX PDGF-immunorelated protein that may play a significant role in the

XX normal development, growth and repair of human tissue and probably

XX functions as a growth factor in wound healing. CTGF may be involved

XX in diseases in which there is an overgrowth of connective tissue

XX cells, such as cancer, tumour formation and growth, fibrotic

XX diseases (e.g. pulmonary fibrosis, kidney fibrosis, glaucoma) and

XX atherosclerosis. Recombinant CTGF can be produced in transformed

XX host cells utilising a cDNA clone isolated from a HUVEC library.

XX It can be used to accelerate wound healing. CTGF inhibitors can be

XX used to treat atherosclerosis and fibrotic diseases such as

XX scleroderma, arthritis, liver cirrhosis, and scarring.

XX Sequence 349 AA;

Query Match 2.0%; Score 7; DB 18; Length 349;

Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 FVVLAL 10  
 |||||  
 Db 13 fvvllal 19

## RESULT 36

AAW11302  
ID AAW11302 standard; Protein: 349 AA.  
XX  
XX AAW11302;  
XX  
DT 18-MAR-1997 (first entry)  
XX  
DE Connective tissue growth factor.  
XX  
KW Connective tissue growth factor; CTGF; human; connective tissue cell;  
KW proliferative disease; platelet-derived growth factor; PDGF; development;  
KW tissue growth; repair; umbilical vein endothelial cell; HVE cell;  
KW antibody; wound healing; cancer; fibrotic disease; atherosclerosis;  
KW inhibitor; protease degradation; growth factor; therapy.  
XX  
OS Homo sapiens.  
XX  
PN US5585270-A.  
XX  
PD 17-DEC-1996.  
XX  
PF 30-AUG-1991; 91US-0752427.  
XX  
PR 30-AUG-1991; 91US-0752427.  
PR 14-DEC-1993; 93US-0167628.  
PR 10-FEB-1995; 95US-0386680.  
XX  
PA (USF-) UNIV SOUTH FLORIDA.  
PI Bradham DM, Grotendorst GR;  
XX  
DR WPI: 1997-051180/05.  
DR N-PSDB: AAT51234.  
XX  
PT New nucleic acid encoding connective tissue growth factor - useful  
PT for accelerating wound healing, also for diagnosis and treatment of  
PT proliferative disease  
XX  
PS Claim 9; Column 15-18; 11pp; English.  
XX  
XX This sequence represents the human connective tissue growth factor  
CC (CTGF). CTGF is related immunologically and biologically to  
CC platelet-derived growth factor (PDGF), but is encoded by an unrelated  
CC gene. CTGF is thought to play a significant role in the normal  
CC development, growth, and repair of human tissue, similarly to PDGF. The  
CC cDNA encoding this sequence was isolated by screening a cDNA library from  
CC human umbilical vein endothelial (HVE) cells with anti-PDGF antibodies.  
CC CTGF can be used to accelerate wound healing. Also, elevated levels of  
CC CTGF may be diagnostic of proliferative diseases involving outgrowth of  
CC connective tissue cells, such as cancer, fibrotic disease and  
CC atherosclerosis. All of these diseases can be treated with reagents  
CC reactive with CTGF, such as antibodies (which can also serve as assay  
CC reagents). Antisense nucleic acids, and ribozymes could also be used to  
CC inhibit CTGF production. The advantage with using CTGF is that it is  
CC more stable, and less susceptible to protease degradation than PDGF, and  
CC other growth factors involved in wound healing. This is believed to be  
CC due to the high Cys content.  
XX  
SQ Sequence 349 AA:

Query Match 2.0%; Score 7; DB 18; Length 349;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 FVVLAL 10  
|||||||  
DB 13 fvvllal 19

RESULT 37

## AAW62084

AAW62084 standard; Protein: 349 AA.  
XX  
XX AAW62084;  
XX  
DT 15-SEP-1998 (first entry)  
XX  
DE Human connective tissue growth factor.  
XX  
KW Human; connective tissue growth factor; CTGF; diagnosis; cancer;  
KW platelet derived growth factor; ameliorating cell proliferative disorder;  
KW atherosclerosis; fibrotic disease.  
XX  
OS Homo sapiens.  
XX  
PN US5783187-A.  
XX  
PD 21-JUL-1998.  
XX  
PF 11-SEP-1996; 96US-0712302.  
XX  
PR 30-AUG-1991; 91US-0752427.  
PR 14-DEC-1993; 93US-0167628.  
PR 11-SEP-1996; 96US-0712302.  
XX  
PA (USF-) UNIV SOUTH FLORIDA.  
PI Bradham DM, Grotendorst GR;  
XX  
DR WPI: 1998-426958/36.  
DR N-PSDB: AAV38085.  
XX  
PT Ameliorating cell proliferative disorder associated with connective  
PT tissue growth factor - comprises the administration of an antibody  
PT which binds to connective tissue growth factor and not to  
PT platelet-derived growth factor  
XX  
PS Example 6; Column 17-20; 11pp; English.  
XX  
XX A method has been developed for ameliorating a cell proliferative  
CC disorder associated with connective tissue growth factor (CTGF). The  
CC method comprises the administration of an antibody or its fragment that  
CC binds to CTGF and not to platelet-derived growth factor (PDGF), to the  
CC site of the disorder. CTGF is related immunologically and biologically  
CC to PDGF. The present sequence represents CTGF. The method is used to  
CC treat conditions involving the overgrowth of connective tissue cells  
CC such as cancer, atherosclerosis and other fibrotic diseases.  
XX  
SQ Sequence 349 AA:

Query Match 2.0%; Score 7; DB 19; Length 349;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 FVVLAL 10  
|||||||  
DB 13 fvvllal 19

## RESULT 38

AAV18361  
ID AAV18361 standard; Protein: 349 AA.  
XX  
XX AAV18361;  
XX

DT 20-AUG-1999 (first entry)  
XX  
DE Human connective tissue growth factor.  
XX  
KW CTGF; connective tissue growth factor; human; fibrotic disease;  
KW cell proliferative disorder; atherosclerosis; diagnosis.  
XX



OS Homo sapiens.  
 XX US5916756-A.  
 XX  
 PD 29-JUN-1999.  
 XX  
 PF 20-JUN-1997; 97US-0880031.  
 XX  
 PR 14-DEC-1993; 93US-0167628.  
 PR 10-FEB-1995; 95US-0386680.  
 PR 11-FEB-1996; 96US-0712302.  
 PR 20-JUN-1997; 97US-0880031.  
 XX  
 PA (UYSF-) UNIV SOUTH FLORIDA.  
 PI Bradham DM, Grotendorst GR.  
 DR WPI: 1999-384720/32.  
 DR N-PSDB: AAX61317.  
 XX  
 PR Detecting cell proliferative disorders such as fibrotic disease and  
 PT atherosclerosis  
 PS Disclosure: Column 15-18; 11pp; English.  
 XX  
 CC This sequence is the human connective tissue growth factor (CTGF).  
 CC The invention relates to a method of detecting a cell proliferative  
 CC disorder comprising comparing the level of CTGF in a sample against a  
 CC control, where an increase is indicative of a cell proliferative disorder  
 CC (fibrotic disease or atherosclerosis). The method is used to detect cell  
 CC proliferative disorders such as fibrotic disease and atherosclerosis.  
 XX  
 SQ Sequence 349 AA;

Query Match 2.0%; Score 7; DB 20; Length 349;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FVVLAL 10  
 |||||  
 Db 13 fvvllal 19

## RESULT 39

AAW81425  
 ID AAW81425 standard; Protein; 349 AA.

AC AAW81425;

DT 25-JAN-1999 (first entry)

DE Connective tissue growth factor (CTGF).

XX CTGF; connective tissue growth factor; bone formation; tissue; arthritis;

KW wound healing; cartilage formation; osteoporosis; osteoarthritis; burn;

KW osteochondritis; skeletal disorder; hypertrophic scar; protease; PDGF;

KW degradation; vascular hypertrophy; platelet derived growth factor.

XX Homo sapiens.

OS Homo sapiens.

XX Key

PD 17-NOV-1998.

PF 31-MAY-1996; 96US-0656393.

XX

PR 31-MAY-1996; 96US-0656393.  
 PR 30-AUG-1991; 91US-0752427.  
 PR 14-DEC-1993; 93US-0167628.  
 PR 10-FEB-1995; 95US-0386680.  
 PR 02-JUN-1995; 95US-0459717.  
 XX  
 PA (UYMI-) UNIV MIAMI.  
 PA (UYSF-) UNIV SOUTH FLORIDA.  
 XX  
 PI Grotendorst GR;  
 DR WPI: 1999-023382/02.  
 DR N-PSDB: AAV65380.  
 XX  
 PR Use of connective tissue growth factor - for inducing bone, tissue  
 PT or cartilage formation in a patient or for inducing wound healing  
 PS Disclosure: Fig 1C; 30pp; English.  
 XX  
 CC This represents the amino acid sequence of connective tissue growth  
 CC factor (CTGF). This can be used in the method of the invention for  
 CC inducing bone or tissue formation that comprises administration to a  
 CC patient, a composition comprising CTGF and a carrier. CTGF can also be  
 CC used in a method for inducing wound healing. The methods can be used for  
 CC inducing bone, tissue or cartilage formation in disorders such as  
 CC osteoporosis, osteoarthritis or osteochondritis, arthritis, skeletal  
 CC disorders, hypertrophic scars, burns, vascular hypertrophy, or in wound  
 CC healing. The CTGF and functional fragments are more stable and less  
 CC susceptible to protease degradation than platelet derived growth factor  
 CC (PDGF) and other growth factors known to be involved in wound healing.  
 XX  
 SQ Sequence 349 AA;

Query Match 2.0%; Score 7; DB 20; Length 349;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FVVLAL 10  
 |||||  
 Db 13 fvvllal 19

## RESULT 40

AAV92939  
 ID AAV92939 standard; Protein; 349 AA.

AC AAV92939;

DT 08-NOV-2000 (first entry)

DE Human connective tissue growth factor protein.

XX Dermatological; antiarthritic; antiarteriosclerotic; antidiabetic;

KW nephrotropic; ophthalmological; hypotensive; cardiant; tranquilizer;

KW vulnary; antiinflammatory; human; connective tissue growth factor;

KW CTGF; extracellular matrix synthesis; collagen synthesis; antibody;

KW myofibroblast differentiation; antisense; fibroproliferative disease;

XX fibrosis; trauma; cancer; inflammation; diabetes; keloid.

XX Homo sapiens.

OS Homo sapiens.

XX Key

PD 14-DEC-1998; 98US-0112240.

PF 14-DEC-1998; 98US-0112241.

XX (UYMI-) UNIV MIAMI.

XX

PI Groendorst GR;  
 XX  
 DR WPI: 2000-431565/37.  
 DR N-PSDB: AAA11278.  
 XX  
 PT Fragment of connective tissue growth factor, useful for treating  
 PT fibroproliferative diseases or disorders, including kidney fibrosis,  
 PT scleroderma, arthritis, hypertrophic scarring, atherosclerosis, diabetic  
 PT nephropathy and retinopathy -  
 XX  
 PS Claim 2, 3; Fig 3A-B; 74pp; English.  
 XX  
 CC This sequence represents a human connective tissue growth factor (CTGF)  
 CC polypeptide having the ability to induce extracellular matrix synthesis,  
 CC collagen synthesis and/or myofibroblast differentiation. The invention  
 CC relates to fragments of CTGF, especially those encoded by exons 2 and/or  
 CC 3, which contain the inductive activity. The protein and/or fragments  
 CC can be used to raise antibodies and the coding sequence can be used to  
 CC generate antisense oligonucleotides. The antibody or antisense sequence  
 CC against the CTGF sequence can be used in a method to treat a  
 CC CTGF-associated disease or disorder such as a fibroproliferative disease  
 CC or disorder, especially selected from kidney fibrosis, scleroderma,  
 CC pulmonary fibrosis, liver fibrosis, arthritis, hypertrophic scarring,  
 CC atherosclerosis, diabetic nephropathy and retinopathy, hypertension,  
 CC kidney disorders, angiogenesis-related disorders, skin fibrotic  
 CC disorders, and cardiovascular disorders. The disease or disorder can  
 CC also be selected from acute or repetitive traumas (including surgery or  
 CC radiation therapy, and fibrosis of organs), diseases caused by vascular  
 CC endothelial cell proliferation or migration (including cancers),  
 CC inflammatory bowel disease, Crohn's disease, joint inflammation,  
 CC interstitial disease, dermatological diseases, diabetes, and keloids.  
 CC  
 XX Sequence 349 AA:  
 SQ  
 Query Match 2.0%; Score 7; DB 21; Length 349;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 FVVLAL 10  
 |||||  
 Db 13 fvvllal 19  
 RESULT 41  
 AAY92940  
 ID AAY92940 standard; Protein: 349 AA.  
 XX  
 AC AAY92940;  
 XX  
 DT 08-NOV-2000 (first entry)  
 XX  
 DE Human connective tissue growth factor protein.  
 XX  
 KW Human; connective tissue growth factor; CTGF; mitogen; antibody; liver;  
 KW fibroproliferative disease; fibrosis; kidney; arthritis;  
 KW hypertrophic scarring; atherosclerosis; diabetic nephropathy; retinopathy;  
 KW hypertension; cardiovascular disorder; wound healing; bone repair.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200035939-A2.  
 XX  
 PD 22-JUN-2000.  
 XX  
 PE 14-DEC-1999; 99WO-US29654.  
 XX  
 PR 14-DEC-1998; 98US-0112240.  
 PR 14-DEC-1998; 98US-0112241.  
 XX  
 PA (UVM1-) UNIV MIAMI.  
 PA (FIBR-) FIBROGEN INC.  
 XX

PI Groendorst GR, Neff TB;  
 XX  
 DR WPI: 2000-431568/37.  
 DR N-PSDB: AAA11280.  
 XX  
 PT New fragment of connective tissue growth factor (CTGF) polypeptide  
 PT having mitogenic activity, useful in wound healing, bone and tissue  
 PT repair -  
 XX  
 PS Claim 2, 3; Fig 2A-B; 71pp; English.  
 XX  
 CC This sequence represents a human connective tissue growth factor (CTGF)  
 CC polypeptide which has mitogenic activity. The protein can be used to  
 CC raise antibodies which specifically bind to CTGF and are used to treat  
 CC a CTGF-associated disease or disorder. e.g. a fibroproliferative  
 CC disease/disorder such as kidney fibrosis, scleroderma, pulmonary  
 CC fibrosis, liver fibrosis, arthritis, hypertrophic scarring,  
 CC atherosclerosis, diabetic nephropathy and retinopathy, hypertension,  
 CC kidney disorders, angiogenesis-related disorders, skin fibrotic  
 CC disorders, and cardiovascular disorders. The protein is also useful  
 CC in wound healing, bone and tissue repair.  
 CC  
 XX Sequence 349 AA:  
 SQ  
 Query Match 2.0%; Score 7; DB 21; Length 349;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 FVVLAL 10  
 |||||  
 Db 13 fvvllal 19  
 RESULT 42  
 AAY44755  
 ID AAY44755 standard; protein: 349 AA.  
 XX  
 AC AAY44755;  
 XX  
 DT 04-MAY-2000 (first entry)  
 XX  
 DE Human connective tissue growth factor.  
 XX  
 KW Connective tissue growth factor; CCN growth regulator; angiogenesis;  
 KW antiangiogenic; basic fibroblast growth factor; bFGF; neovascular;  
 KW endothelial cell proliferation; retinal; haemangioma; leukemia;  
 KW metastasis; psoriasis; tumour; glaucoma; diabetic retinopathy; arthritis;  
 KW endometriosis; insulin-like growth factor-binding domain; IGF;  
 KW von Willebrand factor type C repeat; Thrombospondin type I domain;  
 KW C-terminal cysteine knot profile; CTCK-2; human.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200005356-A1.  
 XX  
 PD 03-FEB-2000.  
 XX  
 PE 11-JUN-1999; 99WO-US13338.  
 XX  
 PR 21-JUL-1998; 98US-0119804.  
 XX  
 PA (CHIL-) CHILDRENS MEDICAL CENT.  
 PA  
 PI Folkman J, Lin J;  
 XX  
 DR WPI: 2000-182688/16.  
 XX  
 PT New anti-angiogenic protein containing an IGF binding, Willebrand  
 PT factor type C, thrombospondin type I and cysteine knot domains is  
 PT useful for inhibiting atopic angiogenesis e.g. in solid tumors -  
 XX  
 PS Claim 6; Page 26-27; 30pp; English.

XX The present sequence is human connective tissue growth factor,  
 CC which is a member of GCM growth regulator family. It has antiangiogenic  
 CC activity and is a potent inhibitor of basic fibroblast growth factor  
 CC (bFGF)-stimulated bovine endothelial cell proliferation. It contains  
 CC insulin-like growth factor (IGF)-binding domain, von Willebrand factor  
 CC type C repeat, Thrombospondin type 1 domain and C-terminal cysteine knot  
 CC profile (CTGF)-2 domain. It can be used to treat conditions associated  
 CC with abnormal angiogenesis or neovascularisation like, retinal  
 CC neovascularisation, tumour growth, haemangioma, solid tumours, leukaemia,  
 CC metastasis, psoriasis, neovascular glaucoma, diabetic retinopathy,  
 CC arthritis, endometriosis and premature retinopathy.  
 CC  
 SQ Sequence 349 AA:

Query Match 2.0%; Score 7; DB 21; Length 349;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 FVLLAL 10  
 |||||  
 DB 13 fvvllal 19

## RESULT 43

AAB60664  
 ID AAB60664 standard; Protein: 349 AA.

AC AAB60664;

DT 22-MAY-2001 (first entry)

DE Human connective tissue growth factor (CTGF).

KW Human CTGF; connective tissue growth factor; recombinant production;

KW mitogenic; chemotactic; tissue development; growth; repair;

KW wound healing; vulnary; diagnostic agent;

XX cellular proliferation disorder.

OS Homo sapiens.

XX

PN US6190884-B1.

PD 20-FEB-2001.

PF 18-MAY-1998; 98US-0080715.

XX

PR 14-DEC-1993; 93US-0167628.

PR 10-FEB-1995; 95US-0386680.

PR 30-AUG-1991; 91US-0752427.

PR 11-SEP-1996; 96US-0712302.

XX

PA (UYSF-) UNIV SOUTH FLORIDA.

XX

PI Grotendorst GR, Bradham DM;

XX

DR MPI: 2001-210379/21.

XX

N-PSDB; AAF5954.

PT Producing connective tissue growth factor involves transforming a host  
 PT cell with polynucleotide encoding the growth factor and growing the  
 PT cell under optimum conditions so that the polynucleotide is expressed  
 CC  
 PS Claim 1; Column 17-20; 11pp; English.  
 CC The invention relates to a method for the recombinant production of  
 CC human connective tissue growth factor (CTGF; AAB60664), involving  
 CC transforming a prokaryotic or eukaryotic host cell with an expression  
 CC construct comprising the CTGF cDNA sequence (AAF5954) or a fragment  
 CC thereof, and culturing the host cell under conditions suitable for the  
 CC expression of CTGF. CTGF is a mitogen and chemotactic agent for

CC connective tissue cells and plays a significant role in normal  
 CC development, growth and repair of human tissues. It is useful as a  
 CC therapeutic for accelerating wound healing and promoting normal healing  
 CC mechanisms and may therefore be used in the treatment of e.g., burns.  
 CC CTGF is also useful as a diagnostic reagent for diagnosing pathological  
 CC states in a patient suspected of having a disease characterised by a  
 CC disorder of cellular proliferation. The present sequence represents  
 CC human CTGF.  
 CC  
 SQ Sequence 349 AA:

Query Match 2.0%; Score 7; DB 22; Length 349;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 FVLLAL 10  
 |||||  
 DB 13 fvvllal 19

## RESULT 44

AAB48831  
 ID AAB48831 standard; Protein: 349 AA.

AC AAB48831;

DT 09-MAR-2001 (first entry)

DE Human connective tissue growth factor (CTGF).

KW Human; connective tissue growth factor; CTGF; PDGF-like activity;

KW platelet-derived growth factor; transforming growth factor-beta;

KW TGF-beta; mitogenic; proliferative; chemotactic; wound healing;

XX vulnary; endothelial cell; fibroblast.

OS Homo sapiens.

XX

PN US6149916-A.

PD 21-NOV-2000.

PF 12-JUN-1998; 98US-0097179.

XX

PR 14-DEC-1993; 93US-0167628.

PR 10-FEB-1995; 95US-0386680.

PR 11-SEP-1996; 96US-0712302.

PR 30-AUG-1991; 91US-0752427.

XX

PA (UYSF-) UNIV SOUTH FLORIDA.

XX

PI Bradham DM, Grotendorst GR;

XX

DR MPI: 2001-079389/09.

XX

N-PSDB; AAC87517.

PT Accelerating wound healing or stimulating growth of connective tissue  
 PT cells involves contacting the site of a wound or cells with a  
 PT composition comprising purified connective tissue growth factor and  
 PT transforming growth factor beta -  
 CC  
 PS Claim 1; Column 17-20; 11pp; English.  
 CC The invention relates to methods of accelerating wound healing in a  
 CC patient. One method involves contacting the site of the wound with a  
 CC composition comprising purified connective tissue growth factor  
 CC (CTGF; AAB48831) and transforming growth factor beta (TGF-beta), which  
 CC stimulates the production of CTGF in vivo. Another method involves  
 CC stimulating the growth of connective tissue cells by contacting the  
 CC cells with CTGF or an active fragment thereof. CTGF is produced by  
 CC endothelial and fibroblastic cells, both of which are present at the  
 CC site of a wound, and is mitogenic and chemotactic for connective tissue  
 CC cells. It has biologically similar activity to PDGF (platelet-derived

CC growth factor), and is also immunologically related to it, but it is  
CC the product of a distinct gene. CTGF is useful for accelerating wound  
CC healing by stimulating the growth of connective tissue cells. CTGF, or  
CC its functional fragments, is more stable and less susceptible to  
CC protease degradation than PDGF and other growth factors known to be  
CC involved in wound healing. The present sequence represents human CTGF.  
XX  
SQ Sequence 349 AA:

Query Match 2.0%; Score 7; DB 22; Length 349;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 FVLLAL 10  
|||||  
DB 13 fvllal 19

## RESULT 45

AAB76742  
ID AAB76742 standard; Protein; 359 AA.

XX AAB76742;

XX 11-APR-2001 (first entry)

XX Corynebacterium glutamicum MCT protein SEQ ID NO:466.

KW Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;  
KW membrane construction and membrane transport protein; petroleum spill;  
KW hydrocarbon degradation; gram positive aerobic bacterium; marker;  
KW identification; microorganism; fine chemical production; transformation;  
KW genome mapping; genetic engineering.  
XX

OS Corynebacterium glutamicum.

PN W0200100805-A2.

PD 04-JAN-2001.

PF 23-JUN-2000; 2000MO-IB00926.

XX 25-JUN-1999; 99US-0141031.  
PR 08-JUL-1999; 99DE-1031454.  
PR 08-JUL-1999; 99DE-1031478.  
PR 08-JUL-1999; 99DE-1031563.  
PR 09-JUL-1999; 99DE-1032122.  
PR 09-JUL-1999; 99DE-1032124.  
PR 09-JUL-1999; 99DE-1032125.  
PR 09-JUL-1999; 99DE-1032128.  
PR 09-JUL-1999; 99DE-1032180.  
PR 09-JUL-1999; 99DE-1032182.  
PR 09-JUL-1999; 99DE-1032190.  
PR 09-JUL-1999; 99DE-1032191.  
PR 09-JUL-1999; 99DE-1032209.  
PR 09-JUL-1999; 99DE-1032212.  
PR 09-JUL-1999; 99DE-1032227.  
PR 09-JUL-1999; 99DE-1032228.  
PR 09-JUL-1999; 99DE-1032230.  
PR 09-JUL-1999; 99DE-1032927.  
PR 14-JUL-1999; 99DE-1033005.  
PR 14-JUL-1999; 99DE-1033006.  
PR 14-JUL-1999; 99DE-1040764.  
PR 27-AUG-1999; 99DE-1040765.  
PR 27-AUG-1999; 99DE-1040830.  
PR 27-AUG-1999; 99DE-1040831.  
PR 27-AUG-1999; 99DE-1040832.  
PR 27-AUG-1999; 99DE-1040833.  
PR 31-AUG-1999; 99DE-1041378.  
PR 31-AUG-1999; 99DE-1041379.

PR 31-AUG-1999; 99DE-1041395.  
PR 03-SEP-1999; 99DE-1042077.  
PR 03-SEP-1999; 99DE-1042078.  
PR 03-SEP-1999; 99DE-1042079.  
PR 03-SEP-1999; 99DE-1042088.  
XX

PA (BAD1) BASF AG.

XX Pompejus M, Kroege B, Schroeder H, Zelder O, Habernauer G;

XX WPI: 2001-071486/08.

DR N-PSDB: AAF67975.

PT Corynebacterium glutamicum nucleic acids encoding membrane construction  
PT and membrane transport proteins or their portions, useful for typing or  
PT identifying C. glutamicum or related bacteria, and as markers for  
PT transformation -  
XX

PS Claim 20; Page 814-815; 1119pp; English.

XX AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane  
CC construction and membrane transport (MCT) proteins given in AAB76510 to  
CC AAB76847. The MCT nucleic acids and proteins are useful in the  
CC identification of microorganisms which can be used to produce fine  
CC chemicals, for modulating fine chemical production in C. glutamicum or  
CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or  
CC identification of C. glutamicum or related bacteria, as reference points  
CC for mapping C. glutamicum genome, and as markers for transformation.  
CC AAF68082 and AAF68082 represent sequencing primers which are used in an  
CC example from the present invention.  
XX

SQ Sequence 359 AA:

Query Match 2.0%; Score 7; DB 22; Length 359;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 VLALVA 12  
|||||

DB 207 vlalva 213

## RESULT 46

AAG44424  
ID AAG44424 standard; Protein; 371 AA.

XX AAG44424;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 55646.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126284.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.

PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134376.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139753.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145228.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147132.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	13-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.

PR 13-OCT-1999; 99US-0159235.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159337.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 2.0%; Score 7; DB 21; Length 371;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 163 SVLSELP 169  
Db 327 svlslp 333  
  
RESULT 47  
AAG50200  
ID AAG50200 standard; Protein: 371 AA.  
XX  
AC AAG50200;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63588.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0123788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132464.  
PR 05-MAY-1999; 99US-0132465.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136382.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144335.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149358.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0159294.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.

PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 2.0%; Score 7; DB 21; Length 371;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 SVLSSLP 169  
DB 327 svlsslp 333

RESULT 48  
AAB58280  
ID AAB58280 standard; Protein: 376 AA.

AC AAB58280;  
XX 14-MAR-2001 (first entry)

DE Lung cancer associated polypeptide sequence SEQ ID 618.

XX Human; lung cancer associated protein; neuroprotective; cytostatic;  
KW Cardioactive; Immunomodulatory; muscular active; vulnery;  
KW gastrointestinal; nephrotropic; antileukemic; gynecological;  
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;  
KW proliferative disorder; wound healing; infectious disease.

XX Homo sapiens.

OS Homo sapiens.  
XX WO200005180-A2.

PN 21-SEP-2000.

PD 08-MAR-2000; 2000WO-US05918.

PF 12-MAR-1999; 99US-0124270.

PR (HUMA-) HUMAN GENOME SCL INC.  
PA (ROSE/) ROSEN C A.

PI Ruben SM;  
XX WPI: 2000-587514/55.  
DR N-PSDB: AAF18156.

XX Lung cancer associated gene sequences, referred to as lung cancer  
PT antigens, useful for treatment, prevention, and diagnosis of disorders  
PT such as lung cancer -

XX Claim 11; Page 1112-1113; 1425pp; English.  
PS Polynucleotide sequences AAF1982 - AAF18424 encode human lung cancer  
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer  
CC associated proteins and polynucleotide sequences, their agonists, and  
CC antagonists may have neuroprotective; cyostatic; cardioactive;

CC immunomodulatory; muscular active general; vulnery; gastrointestinal  
 CC general; nephrotropic; antineoplastic; gynecological; or antibacterial  
 CC activity. The invention also includes antibodies specific for the  
 CC protein or polynucleotide sequences. The lung cancer associated  
 CC polynucleotide sequences may be used for detection of lung cancer,  
 CC chromosome identification, as chromosome markers, and for numerous other  
 CC diagnostic or research purposes. The proteins may be used to treat  
 CC disorders such as neural, immune, muscular, reproductive,  
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
 CC disorders. The proteins may also be used in the treatment of wounds and  
 CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and  
 CC peptide AAB58549 are used in the course of the invention for the  
 CC identification and characterization of the polynucleotide and protein  
 CC sequences.

CC Sequence 376 AA:

Query Match 2.0%; Score 7; DB 21; Length 376;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 43 PDVAALS 49  
 |||||  
 Db 263 pdvaals 269

RESULT 49

AAW21965  
 ID AAW21965 standard; Protein: 377 AA.

AC AAW21965;

DT 02-DEC-1997 (first entry)

DE Human cyclin I.

KM human; cyclin I; antisense; probe; neurone; cancer; antibody.

OS Homo sapiens.

PN W09712973-A1.

PD 10-APR-1997.

PF 07-OCT-1996; 96WO-JP02905.

PR 05-OCT-1995; 95JP-0284663.

PA (SOME ) SUMITOMO ELECTRIC IND CO.

PI Nakamura T;

XX WPI: 1997-226217/20.

DR N-PSDB: AAT73937.

PT Human cyclin I protein and related (anti:sense) DNA - used for  
 neuron labelling method and cancer cell detection

PS Claim 1; Fig 1; 45pp; Japanese.

CC This sequence is human cyclin I. Antisense polynucleotides are  
 CC useful for as probes and can be labelled and used for detection of  
 CC neurones by hybridisation with mRNA for cyclin I (contained in the  
 CC neurones and arising by the expression of the cyclin I gene in these  
 CC cells). The gene can be used for detection of cancer cells by detecting  
 CC the expression of the cyclin I gene in these cells. Also antibodies  
 CC specific for the fragments of the protein (especially AAW21966) can be  
 CC used for detection.

CC Sequence 377 AA:

Query Match 2.0%; Score 7; DB 18; Length 377;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 166 SSPLNS 172  
 |||||  
 Db 252 ssplns 258

RESULT 50

AAW52185  
 ID AAW52185 standard; Protein: 377 AA.

AC AAW52185;

DT 09-FEB-2000 (first entry)

DE Human cyclin I amino acid sequence.

KW Cyclin-dependent kinase 2; CDK2; hsrq; hsrq+2; cyclin I;  
 ERH; cell cycle; proliferation; cancer; hyperproliferative disorder;  
 atherosclerosis; tumour.

OS Homo sapiens.

PN W09925829-A2.

PD 27-MAY-1999.

PF 12-NOV-1998; 98WO-US24095.

PR 13-NOV-1997; 97US-0969106.

PA (CURA-) CURAGEN CORP.

PI Yang M, Nandabalan K, Schulz VP;

XX WPI: 2000-061923/05.

DR N-PSDB: AAZ37836.

PT New complexes of the cyclin-dependent kinase 2 protein with its  
 interacting proteins, used to treat, e.g. atherosclerosis -

PS Examples; Fig 2; 90pp; English.

CC This is the human Cyclin I amino acid sequence. Cyclin I is expressed at  
 CC almost constant levels throughout the cell cycle, and is implicated in  
 CC controlling cell cycle progression and transcriptional control. Cyclins  
 CC form complexes with cyclin-dependent kinases. CDK2, cyclin-dependent  
 CC kinase 2, is used in the invention which relates to complexes of the  
 CC CDK2 protein with other proteins, selected from cyclin I, ERH, hsrq,  
 CC hsrq+1 and hsrq+2 (AAW52185-Y52188). CDK2 is expressed late in G1 or  
 CC early in S phase of the cell cycle, and is pivotal for G1/S transition.  
 CC Compositions containing a CDK2 complex, an antibody targeting the  
 CC complex, and nucleotide sequences encoding CDK2 or its derivatives can  
 CC be used therapeutically. The complexes and their nucleotide sequences can  
 CC be used to treat diseases or disorders associated with increased or  
 CC decreased levels of the complex. Screening the complex, or a derivative  
 CC or a modulator of the complex for neoplastic activity by measuring the  
 CC survival or proliferation of cells from a malignant cell line when in  
 CC contact with the complex can be used to indicate if the the complex has  
 CC anti-neoplastic activity. Screening for molecules that modulate the  
 CC formation of the complexes can be used for treating or preventing  
 CC atherosclerosis or atherosclerosis-associated disease by contacting  
 CC cells or using a test animal, in which tumour growth or regression is  
 CC measured to test whether anti-neoplastic activity is displayed. Diseases  
 CC which can be treated or prevented by molecule/s which modulate the  
 CC function of the complex include cancer, hyperproliferative disorders and  
 CC atherosclerosis.

CC Sequence 377 AA:



Mon Aug 20 10:21:48 2001

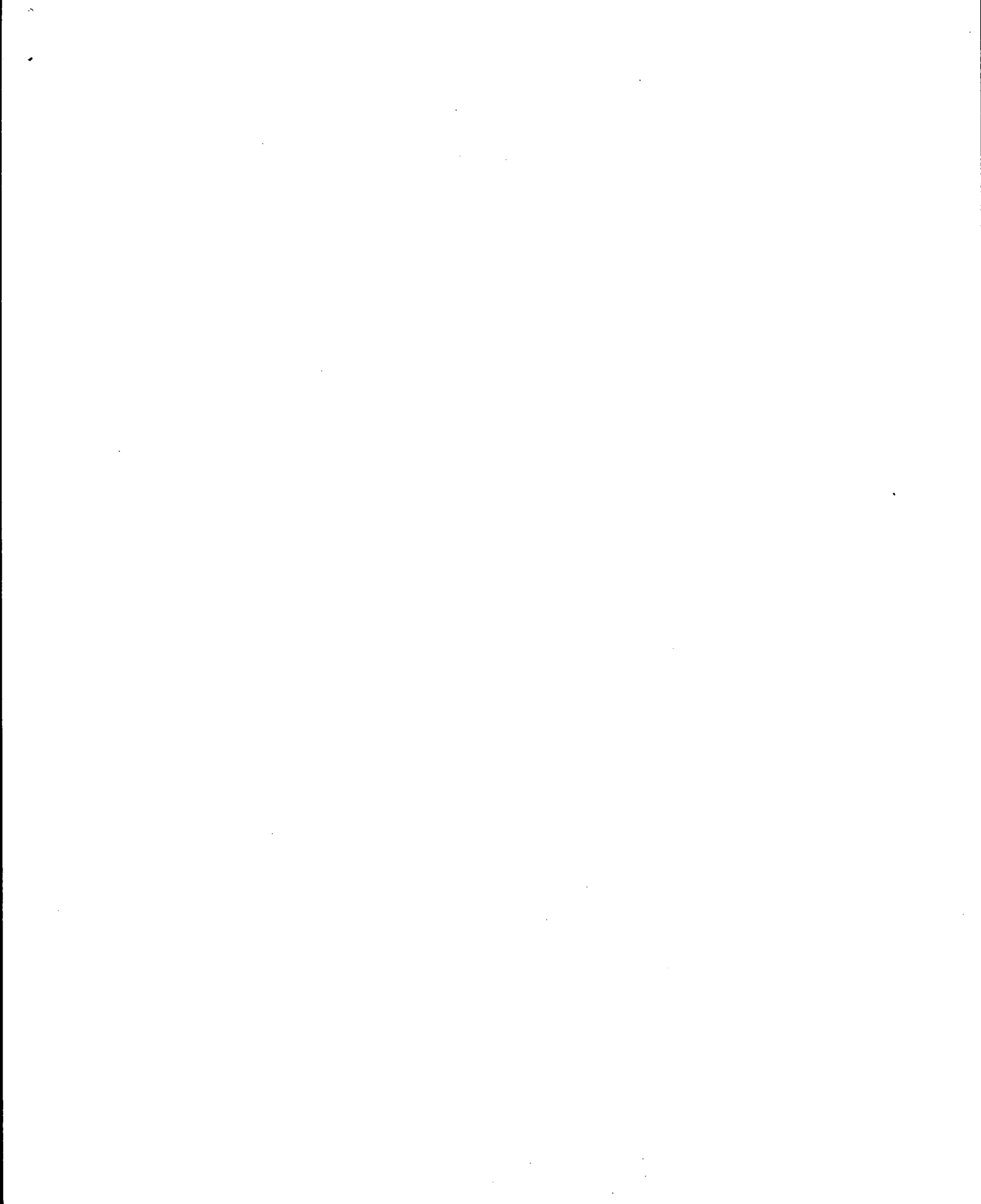
us-09-284-320-6.oli5.rag

Page 43

Query Match 2.0%; Score 7; DB 21; Length 377;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0;  
Gaps 0;  
Qy 166 SSPLNS 172  
|||||||  
Db 252 ssiplns 258

Search completed: August 17, 2001, 16:45:30  
Job time: 132 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 18, 2001, 20:15:30 ; Search time 3117.45 Seconds  
(without alignments)  
6164.542 Million cell updates/sec

Title: US-09-284-320-56

Perfect score: 2033  
Sequence: 1 gagtcgagcgcgctcacctc.....gttaatgatgatccaac 2033

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

```
EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*
```

```
117: qb_est148:*
118: qb_est149:*
119: qb_est150:*
120: qb_est151:*
121: qb_est152:*
122: qb_est153:*
123: qb_est154:*
124: qb_est155:*
125: qb_est156:*
126: qb_est157:*
127: qb_est158:*
128: qb_est159:*
129: qb_est160:*
130: qb_est161:*
131: qb_est162:*
132: qb_est163:*
133: qb_est164:*
134: qb_est165:*
135: qb_est166:*
136: qb_est167:*
137: qb_est168:*
138: qb_est169:*
139: qb_est170:*
140: qb_est171:*
141: qb_est172:*
142: qb_est173:*
143: qb_est174:*
144: qb_est175:*
145: qb_est176:*
146: qb_est177:*
147: qb_est178:*
148: qb_est179:*
149: qb_est180:*
150: qb_est181:*
151: qb_est182:*
152: qb_est183:*
153: qb_est184:*
154: qb_est185:*
155: qb_est186:*
156: qb_est187:*
157: qb_est188:*
158: qb_est189:*
159: qb_est190:*
160: qb_est191:*
161: qb_est192:*
162: qb_est193:*
163: qb_est194:*
164: qb_est195:*
165: qb_est196:*
166: qb_est197:*
167: qb_est198:*
168: qb_est199:*
169: qb_est200:*
170: qb_est201:*
171: qb_est202:*
172: qb_est203:*
173: qb_est204:*
174: qb_est205:*
175: qb_est206:*
176: qb_est207:*
177: qb_est208:*
178: qb_est209:*
179: qb_est210:*
180: qb_est211:*
181: qb_est212:*
182: qb_est213:*
183: qb_est214:*
184: qb_est215:*
185: qb_est216:*
186: qb_est217:*
187: qb_est218:*
188: qb_est219:*
189: qb_est220:*

190: qb_est221:*
191: qb_est222:*
192: qb_est223:*
193: qb_est224:*
194: qb_est225:*
195: qb_est226:*
196: qb_est227:*
197: qb_est228:*
198: qb_est229:*
199: qb_est230:*
200: qb_est231:*
201: qb_est232:*
202: qb_est233:*
203: qb_est234:*
204: qb_est235:*
205: qb_est236:*
206: qb_est237:*
207: qb_est238:*
208: qb_est239:*
209: qb_est240:*
210: qb_est241:*
211: qb_est242:*
212: qb_est243:*
213: qb_est244:*
214: qb_est245:*
215: qb_est246:*
216: qb_est247:*
217: qb_est248:*
218: qb_est249:*
219: qb_est250:*
220: qb_est251:*
221: qb_est252:*
222: qb_est253:*
223: qb_est254:*
224: qb_est255:*
225: qb_est256:*
226: qb_est257:*
227: qb_est258:*
228: qb_est259:*
229: qb_est260:*
230: qb_est261:*
231: qb_est262:*
232: qb_est263:*
233: qb_est264:*
234: qb_est265:*
235: qb_est266:*
236: qb_est267:*
237: qb_est268:*
238: qb_est269:*
239: qb_est270:*
240: qb_est271:*
241: qb_est272:*
242: qb_est273:*
243: qb_est274:*
244: qb_est275:*
245: qb_est276:*
246: qb_est277:*
247: qb_est278:*
248: qb_est279:*
249: qb_est280:*
250: qb_est281:*
251: qb_est282:*
252: qb_est283:*
253: qb_est284:*
254: qb_est285:*
255: qb_est286:*
256: qb_est287:*
257: qb_est288:*
258: qb_est289:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

JOURNAL REFERENCE AUTHORS	Methods Enzymol. 303, 19-44 (1999)
TITLE	2 (sites)
JOURNAL MEDLINE REFERENCE AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, H., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)
TITLE	3 (sites)
JOURNAL MEDLINE REFERENCE AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishikawa, K., Katsunuma, T., Tashiro, H., Itoh, M., Kikuchi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishibe, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasahigashi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, M., Yonekura, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsushima, S., Okazaki, Y., Muramatsu, M., Inoue, Y. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system-384-Format sequencing pipeline with 384 multiplexed capillary sequencer Genome Res. 10 (11), 1757-1771 (2000)
TITLE	4 (sites)
JOURNAL MEDLINE REFERENCE AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)
TITLE	5 (bases 1 to 1989)
JOURNAL REFERENCE AUTHORS	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Horii, F., Imoto, K., Ishii, Y., Itoh, M., Izawa, M., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishikawa, K., Nomura, K., Numasaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shihagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y. Direct Submission Submitted (10-JUL-2000) yoshinide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
TITLE	Further details.
JOURNAL	CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer

COMMENT

with BamHI and XhoI. Vector: a modified pRiEscrept AS(+) altered bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.

## FEATURES

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGI:1902866"
/db_xref="MGI:1917745"
/clone_id="5730403E06"
/clone_1b="RIKEN full-length enriched mouse cDNA library"
/dev_stage="8 days embryo"
CDS
823..1122
/note="putative"

```

```

/codon_start=1
/protein_id="PAB30766.1"
/db_xref="GI:12856740"
/translation="MKSIVGKANNVVELVYKSEPTLSVRSKSTILEAKOENTOSPYNL
AKRYNLEISVYFNLVIMIGIALAVLITTSINIMNDPQDILYRMNOKIRID"
polyA_signal
1973..1978
/note="putative"
1989
polyA_site
/note="putative"
BASE COUNT      531 a      371 c      451 g      636 t
ORIGIN

```

```

Query Match      59.2%, Score 1203.8; DB 192; Length 1989;
Best Local Similarity 79.2%; Pred. No. 6;6e-286;
Matches 1592; Conservative 0; Mismatches 377; Indels 42; Gaps 12;

```

```

29 gggcgctgccccgtgccccgcggccgtccgtggtgccccgcagtgctggccgc 88
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  2 GCAGTGCCTGCTGAGATCCCGGCTCCCTTTTGATGCTCATCTCCGCTTCAGCAAC 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  89 gggcaccatgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 145
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  62 GTGGCACCACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 121
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  146 acgagttatataataatcaccagagctgtgtgtgtgtgtgtgtgtgtgtgtgt 205
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  122 ATGAATTTACATATTAGATCGCCAGAGCTGTGTGTGTGTGTGTGTGTGTGTGT 181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  206 taccagaagaagcgatccacagagtgatgtgtgtgtgtgtgtgtgtgtgtgtgt 265
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  182 TACACAGAGATCGAATCCACAGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTG 241
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  266 acccttctgcccagagctcgaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 325
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  242 ACCTGCTTGGCAGAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 301
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  326 tggatgagtgagagagtgagagagtgagagagtgagagagtgagagagtgagag 385
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  302 TGGGATGATGAGAGGGGGGTGATTAACCTGACCTTCGCGAGGAGCTGATCTAC 361
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  386 ctgtgagaatgcaatctccttcttcttcttcttcttcttcttcttcttcttct 445
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  362 CTGTGAGATGCACTGCTCTTTAGCTAGACAGGCTTGCAAAATGCAATTCCTTAT 421
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  446 tttctgagaagaatccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 505
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  422 TTTGCGAGGAAACACCTGTAGTTTTCAGTTGCTGCCAGCGAGAGAGATATATG 481
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  506 taggaagaagcaactcaatggttgaagaaccttcaagcttcaagcttcaagcttca 565
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  482 TGGGAAAGCAAACTCGTGTGTAAGACCTTTCAGTACATTTGCCGAGCTCCGTAA 541
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  566 gctgtgttgaagaagaactcgttctcaatgtcaatgtcaatgtcaatgtcaatgt 625
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  542 GCTGTTTCAAAAACCTCCCTTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 601
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  626 atgaagttagacctgtcttcttcttcttcttcttcttcttcttcttcttctt 685
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  602 ATGAGGTAGCACTGCTCTTCTCTCCGAACTGCAAGTCACTGATGATTTTCACT 661
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  686 tgcctgtatagaatcagctcgaagaagcagctcagctcagctcagctcagctcag 745
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  662 TGCTGTGATATAGATCTTAGGCAAGGACCATTCACCGACTTGTATCTATGAGCTG 721
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  746 caagtttagatgaatitggagagcgctatggagagagagcttgaacatcagagag 805
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  722 CAGGTTTGGATGAATCTGGGAGCGTTATGGGGAAGACTGTGAACGTTGACGAGTCT 781
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  806 ctgaagcttctgttgaagctgtcgaagaagtttgaagatgaagatgaagatgaag 865
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  782 CTAAAGATCTCTGTGTGATGCTCCCAAAAGTTTGACAGATGACATGATCTCTAT 841
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY      866 ggaatgcagtgatgagttgagctgcgtcgaagctatgtgacacctccatcagaaga 925
DB      842 GGAACGGAGAGGTAGAGTATGAGCTCAAACTGACACATCCCTTGTGAGAGAGT 901
QY      926 caagactacatccttggagcaaaacgaacgagagaccagcaagctccatccttga 965
DB      902 CAAGGACCATCTTGAAGCAAAACAG---AGACACCCAAAGTCCCTATTAACCTTGCAT 958
QY      986 ataagtaatttgaataatccgtgtgtgttcaacatgttgaatgtgataatgcgcct 1045
DB      959 ATAAATATATTTGGAGATTAATCACTGCTTTCACTGCTGATGATGATGAGCT 1018
QY      1046 tggcctgtgtgtatataatcccttcaacatatttgaacatgtgataatataata 1105
DB      1019 TGGCTGCTGCTGATATATCACTCTTACACATTTGGAATATGATCTCGATATGATA 1078
QY      1106 gcatcattatagatgacaaacagaagatcgaatgtgattgtaccctgtgcag 1165
DB      1079 GCATCATCTATAGGATGACAAATCAGAGATTCGAAATGATGAAATTTCTTATGCTAG 1138
QY      1166 aattagaagaaggggttgaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1225
DB      1139 ACTTAGACAAACAGGTTTGTATATGGCTGTTTGTAAAGTATATCTTTAGTGTGTTT 1198
QY      1226 aaagttagatgataatcattacatta-taaaaaaatcoaatltgttcttatttgt 1284
DB      1199 AAAGTACAGGATACCTTAAGTTTATTAAGAAACAAATAGATTTCAATATCTATTTT 1258
QY      1285 gtgtgcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1344
DB      1259 GTGTGCTGCTATGTTGTTCTTATGAGATGATTTATGATGTGAC--AAGTTGAATGTGATA 1316
QY      1345 tagatccatataatgtcttgaatataatgatatagacatlttaacatgttcaatc 1404
DB      1317 TAGATTC---AATGCTCAAAATATGTTGACATAGCCATTTAATATGATTTCAATTC 1373
QY      1405 tgttcaatgaatttgaataatcgaactgaagaagaatgttaaacatttgaatgaatcgtgt 1464
DB      1374 CATTTAATACATGTGGAATATGCACTGAGAGAACTGTAAACCATGTGATGATCTC 1433
QY      1465 ttatgaaagaagtgacatgaattatlttagacaacttcaagatgtcctaactttaa 1524
DB      1434 GCGCTGGAAGAA--CGCACTGCGGTTCTTGAAGAACCTTA-AAATGCTGACCTTCTTCAC 1490
QY      1525 cagcataggtgaaatcataatlttggcctatgttactatgaacaattgtcaatgtcct 1584
DB      1491 AGGCT-----ACGTAGTCACACTGCTTATGACATGAACTATTTGTAATGTCTT 1540
QY      1585 aatttgaatgaataactctgaaacagaagaagaagltttaaacttagatagcctaag 1644
DB      1541 GATCTGCTGTAAATATCTCTGAGACGAGGAAATATGTAACATAAGAGCCCAAAA 1600
QY      1645 atatgagatgtctatataatcgtctagttttagaactgtatcagtaacaagaagacag 1704
DB      1601 TACTGATATGCTTATTCATTCGATCTGATGCTGAGAGCTGATATGATGACGA---AGCT 1656
QY      1705 ctgtttttaaacctcttctgcaagttgtgtgacctacatgtggcctaataatga----- 1757
DB      1657 AGCTGTTTATATCTCTTCTGCAAGTTGTGATCGGTGCTTATGAGAGATTAATAA 1716
QY      1758 --tactaaanaactaataatgtatclaaagaagaactagcctgtgtgaatcl--atagatc 1813
DB      1717 ATTAATAAATAATACTATGTTGATTCAGAGACAGACTAGCTGTTGTGATATAGATCTTG 1776
QY      1814 ctcttcaatataacacaaaatccctgaaggaactttagagcatgtatataaacatc 1873
DB      1777 TAATTAAGTATATAAACAATAATCTGTGGGATATCTTGAAGCATGAAATATAAATAAT 1836
QY      1874 ttatattcaagtaactttcccccgt--tgtaagtactaagttgtgtgtaacatctcat 1931
DB      1837 TTTATTTCTGTACTTCCCTGTAAGATTAATCTGTTGTTGTGATGATGATGCTTCA 1896
QY      1932 tctatagaatatagtgagaagtgagtgaaatcttacttlttatgttgaagtgacaatgt 1991

```







Db	220	GTTTTCGAAATGAAATGGACTTATACAGAGAGACGGATCCACAGCTGGCTGCATTG	279
Qy	241	tcgaatggcttcctcgtgaaagaagacttctctggcaggagccag- tgggttaacct	299
Db	280	TCGATGGGCTTCCTGTGGAAGAGACCTTTCTTGGCCAGGACTCGAGCTGGTAACT	339
Qy	300	gtttacatctcctgggctacgcatacgtatgltgaaagatgaaacaacttgctct	359
Db	340	GTTCATCTCTCCCTCGGGCTACCTCATGTGTATGTGAAGGAGTGAACAACATGGCTCT	399
Qy	360	accccaggcagtgatcattcttcgacaccttgagaaatgacgtccctttagctctgaacg	419
Db	400	ACCCGAGGAGATGTCAATTTCGTACACCTTTGGAGAAATGACATGCTCTTTAGCTTGACAG	459
Qy	420	tgttgcaaatccaatccaatccctatcttctggagaaactcgtgttcttccaatggc	479
Db	460	TGTTGCAAAATTCATTCACCTCCTTATTTCGTAGAGAAACTCTGTGTGTTGCAATTGGC	519
Qy	480	tcccagtgagagaa- agatgtlatatgtagaagaaagcaactcagtglttgaagacctt	538
Db	520	TCCGACGAGAGAAACAGATGTATATGTATGGAGAGGCAACATCAGTGTGGAAGACTTT	579
Qy	539	cagtcaccttggcgcagctgcgtaaacgc- tgttccaagaacctcgtctccaatca	597
Db	580	CAGTCACCTGGGCGGACAGTCCGTAATCCGCTGTTGTTCAAGAAAATCTGTCTCAGTTCA	639
Qy	598	ctcccccctaattctctgagtagaagaaatgaagtggaccctctcttcttcttgaactg	657
Db	640	CTCCCCCTCAATTCCTCTCAGTAGAGAAACAATAAGTTGACGCTGCTTCTTCTTGACCTG	699
Qy	658	caagtgtcaatgatattcttaagcttgc- tgtccgcgtcaatgaatcttagccaagatca	716
Db	700	GAAAGCTACATGATATATCCAGCTGTTGCTGTCTGCATAAAGCATCTACGCCAAGATCA	759
Qy	717	ttctccgattatataca 735	
Db	760	TTCTCTGATTATATATCA 778	

```

FEATURES             SOURCE
source              1..681
                     Location/Qualifiers
                     1..681
DEFINITION           BG530418      681 bp      mRNA
                     B6255916PRL NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4697177 ,
                     mRNA sequence.
ACCESSION            BG530418
VERSION              BG530418.1   GI:13521955
KEYWORDS             EST.
SOURCE               human.
ORGANISM             Homo sapiens
                     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                     Mammalia; Euarchonta; Primates; Carnivora; Homnidae; Homo.
REFERENCE            1 (bases 1 to 681)
AUTHORS              NIH-MGC http://mml.nhl.nih.gov/.
TITLE                National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL              Unpublished (1999)
COMMENT              Contact: Robert Strausberg, Ph.D.
                     Email: cgapds.remall.nih.gov
                     Tissue Procurement: ATCC
                     CDNA library Preparation: CLONETECH Laboratories, Inc.
                     Genomic Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
                     DNA Sequencing by: Incyte Genomics, Inc.
                     Clone distribution: MGC clone distribution information can be
                     found through the I.M.A.G.E. Consortium/LNLT at:
                     http://image.llnl.gov
                     Plate: LCM1525 row: 1 column: 18
                     High quality sequence stop: 681.

```

```

/tissue_type="embryonal carcinoma"
/lab_host="DH10B (TI phage-resistant)"
/notes="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccgcctgcgc); Site_2: SfiI (ggcccatatgcgc);
Double-stranded cDNA was prepared from cell line RNA. 5'
and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCCATTATGAGGC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCCGAGAGCGGCCGACATG-drr(30)BN-3'.
Average
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
library."
```

Query Match	32.2%	Score 654	DB 154	Length 681
Best Local Similarity	99.2%	Pred. No. 1,7e-150		
Matches 657	Conservative 0	Mismatches 5	Indels 0	Gaps 0

QY	328	gtatgtgaaggaaglyaacaaactgctacaccccaagcagtgatcattctgtacct	387
Db	14	GCCTGCTATGGAGGAGTGAACAAACCTGGCTCTACCCCCAGGACAGTATTGGTACCTT	73
QY	388	ttgagaatgcaggttcctcttagctctgaacagttgcaattcattcacctcattt	447
Db	74	TTGGAGAAATGCAGTTCCTTTTAGTCTTGAACAGTGTTCGAATTCCATTCACTCTATT	133
QY	448	tctgaagaaactcctgttctgttttgcagttgctcccaaglyagaaagagtgatatgta	507
Db	134	TCTGAGAAACCTCGTTGGTGTTCAGATTGGCTCCAGTAGGAAAGGTGTATTAGTA	193
QY	508	gggaaggaacaactcagtgattggaagaccttcagtcaccttgcgcaactccgtaacgc	567
Db	194	GCGAAGGCAAACTCACTAGTGTGAAGACCTTTAGTCACTGGGGCAGCTCCGAAATGCG	253
QY	568	ctgtttcaagaaactctgtctctcagttcaactcccccatactctcagtagaagaacat	627
Db	254	CTGTTTCAGAAACCTCGTTGCTCAGTTCACCTCCCTCAATTCCTGAGTAGGAACAT	313
QY	628	gaagtgaccgtctctctctcgaactgaagtgctacatgatattcaagcttgctg	687
Db	314	GAAGTGACCTGCTCTTCTTTTGAAACGTGAAGGCTACATGATATTTCAGCTGTG	373
QY	688	tctgtcataagcatcagccaagatcaattctcgattatattaccttgagcctgga	747
Db	374	TCTGCTCATAGGATCTTAGCCAAAGATCATTTCTCTGATTAATTCATCTGGACCTGGA	433
QY	748	ggtttggaabaaatlggaagcgtctatbgggaagacactgaabaattcaagatgctct	807
Db	434	GSTTTGGATGAATTTGGGAACGTTATGGGAAAGCTGTGAAACAATTGACAGATGCTCT	493
QY	808	aagatcctgttgtagcgtctgcgaagaatttgtagatgaactgtacaagtcattatg	867
Db	494	AAGATCCCTGTGTGACGCTCTGCCAAATTTGGAAGTGAACATGTACAGTCTTTATGGGG	553
QY	868	aatgagtggtatagcttgaactgcgtcaagtcatttgacacctccctcaattagaagaca	927
Db	554	AATGACGTGGTACAGTTATGCTACTGTCAATGTCATTTGACCTCCCTCATTAGGAAGCA	613
QY	928	aagactatccttgaggcaaaacaggaagaaaccagaagtcctccataaacttgat	987
Db	614	AGGACTATCTTGAGGCAAAACAGGCAAGAACCCAGCAAGTCCTATTAACCTTGCAAT	673
QY	988	aa 989	
Db	674	AA 675	

RESULT 6

BC28867

LOCUS BG288867 665 bp mRNA EST 21-FEB-2001  
 DEFINITION 602383808F1 NIH\_MGC\_93 Homo sapiens cDNA clone IMAGE:4513009 5',  
 mRNA sequence.  
 ACCESSION BG288867  
 VERSION BG288867.1 GI:13044136  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 665)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cga@bbs.femail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLM10398 row: p column: 02  
 High quality sequence stop: 665.  
 FEATURES  
 Source  
 1. 665  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4513009"  
 /clone\_lib="NIH\_MGC\_93"  
 /tissue\_type="transitional cell papilloma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: Bladder; Vector: pCMV-SPORT6; Site: 1: NotI;  
 Site: 2: SalI; Cloned unidirectionally; oligo-dT primed.  
 Average insert size 1.7 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH-MGC library."  
 BASE COUNT 185 a 128 c 145 g 207 t  
 ORIGIN

Query Match 32.0%; Score 649.8; DB 175; Length 665;  
 Best Local Similarity 99.4%; Pred. No. 1.8e-149;  
 Matches 663; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

OY 556 ctccgaatcgctgttcaagaacatctgttctgaatcccccattctctg 615  
 |||||||  
 DB 1 ctccgaatcgctgttcaagaacatctgttctgaatcccccattctctg 60  
 |||||||

OY 616 agtagaacaatgaattgacctgtcttcttctgaactgaagtcacatgatatt 675  
 |||||||  
 DB 61 agtagaacaatgaattgacctgtcttcttctgaactgaagtcacatgatatt 120  
 |||||||

OY 676 tcaagcttgctgtctgaatgaatcgaagaatattcccgattatca 735  
 |||||||  
 DB 121 tcaagcttgctgtctgaatgaatcgaagaatattcccgattatca 180  
 |||||||

OY 736 ctggagctggcaggcttgatgaatctgggaacgctatgggaagactctgaacatc 795  
 |||||||  
 DB 181 ctggagctggcaggcttgatgaatctgggaacgctatgggaagactctgaacatc 240  
 |||||||

OY 796 agagatgcttgaatccttctgaagctctgaacaaatttcagatgaatgaat 855  
 |||||||  
 DB 241 agagatgcttgaatccttctgaagctctgaacaaatttcagatgaatgaat 300  
 |||||||

OY 856 cttatggtgggaatcagtgatgaatctgaatcctgaatcctgaacatc 915  
 |||||||  
 DB 301 cttatggtgggaatcagtgatgaatctgaatcctgaatcctgaacatc 360  
 |||||||

OY 916 attagaagaacaagatccttgaagcaaaagaagaacccaggaatccctat 975  
 |||||||  
 DB 361 attagaagaacaagatccttgaagcaaaagaagaacccaggaatccctat 420  
 |||||||

OY 976 aacctgacataagataaatttgaatattccgtagtttcaacatgtagata 1035  
 |||||||  
 DB 421 AACCTTCATATTAAGTATTAATTTGAAATTTCCGCTGTTTCAACATGCTACTTGGAATA 480  
 |||||||

OY 1036 atgacgaccttgacctgtagatattacacaccttaccataatttggaaatgacct 1095  
 |||||||  
 DB 481 ATGATGCGCTTGCGCTTGCTGATTTACCTCTTTACATTTTGGAAACATGGAATCT 540  
 |||||||

OY 1096 ggaataagaacatcattatagatgacaacacgaagatcgaatgattgaatt 1155  
 |||||||  
 DB 541 GGATATGATAGCATCATTTATAGATGACAAACGAAAGATTCGATGATGATGAT 600  
 |||||||

OY 1156 cctgaccgaatagaaaggaggtggaatgagctgttggtaaatatctttt 1215  
 |||||||  
 DB 601 CCTGTGCGAGATTTGAAAAGGGGCTTGGAAATTTGCTGTTT--GTTAATATATCTTTT 658  
 |||||||

OY 1216 agtgatc 1222  
 |||||||

DB 659 AGTGTGC 665

RESULT 7  
 BG538574 878 bp mRNA EST 03-APR-2001  
 LOCUS BG538574  
 DEFINITION 602567303F1 NIH\_MGC\_77 Homo sapiens cDNA clone IMAGE:4691712 5',  
 mRNA sequence.  
 ACCESSION BG538574  
 VERSION BG538574.1 GI:13530807  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 878)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cga@bbs.femail.nih.gov  
 Tissue Procurement: Clontech Laboratories, Inc.  
 CDNA Library Preparation: Clontech Laboratories, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLM1511 row: f column: 01  
 High quality sequence stop: 615.  
 FEATURES  
 Source  
 1. 878  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4691712"  
 /clone\_lib="NIH\_MGC\_77"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: Lung; Vector: pDNR-LIB (Clontech); Site: 1:  
 SfiI (ggccgcctggcc); Site: 2: SfiI (ggccatataggcc); 5' and  
 3' adaptors were used in cloning as follows: 5' adaptor  
 sequence: 5'-CAGCGCATTTATGGCC-3' and 3' adaptor sequence:  
 5'-ATTCTAGAGCGGAGGCGGCACATG-dT(30)BN-3' (where B = A,  
 C, or G and N = A, C, G, or T). Average insert size 1.9  
 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts  
 by PCR. This library was enriched for full-length clones  
 and was constructed by Clontech Laboratories (Palo Alto,  
 CA). Note: this is a NIH-MGC library."  
 BASE COUNT 278 a 121 c 200 g 279 t  
 ORIGIN

Query Match 31.8%; Score 646.4; DB 155; Length 878;  
 Best Local Similarity 88.7%; Pred. No. 1.3e-148;  
 Matches 745; Conservative 0; Mismatches 91; Indels 4; Gaps 4;



OY 1307 gagtgaatagatagatagacgtgaatccacgtgtgtagatattcaataatgcttga 1366  
 |||||  
 Db 607 GAGTGAATATAGATATGACGTGAAT-CCACGTGTGATTTAT--CCATATATGCTTAA 663  
 |||||

OY 1367 tattatagatagacattataac--attgattcattctcgtttaaattgaatt 1424  
 |||||  
 Db 664 TATTATATATAGACATTTTATACATGATTTTCAATTCGCTATGAAATTTGGAA 723  
 |||||

OY 1425 atgactgaaga 1437  
 |||||

Db 724 TGGCTGAGAAA 736

RESULT 9  
 AV721438 753 bp mRNA EST 16-OCT-2000  
 LOCUS AV721438 HTB Homo sapiens cDNA clone HTBALD05 5', mRNA sequence.  
 DEFINITION AV721438  
 ACCESSION AV721438.1 GI:10818590  
 VERSION EST.  
 KEYWORDS  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 753)  
 AUTHORS Gu, Y., Peng, Y., Song, H., Huang, Q., Yang, Y., Gao, G., Xiao, H., Xu, X.,  
 Li, N., Qian, B., Liu, F., Ou, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu,  
 S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R.,  
 Chen, J., Chen, Z. and Han, Z.  
 TITLE Homo sapiens cDNA HTB clones  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Zengqiang Han  
 Chinese National Human Genome Center at Shanghai  
 351 Guo Shouling Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
 201203, P. R. China  
 Tel: 86-21-50801919 (ex.45)  
 Fax: 86-21-50801922  
 Email: hanzq@chgc.sh.cn

FEATURES  
 Source Location/Qualifiers  
 1..753  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="HTBALD05"  
 /clone\_lib="HTB"  
 /tissue\_type="Hypothalamus"  
 /dev\_stage="Adult"  
 /lab\_host="SOLR"  
 /note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2:  
 XhoI"  
 This clone is available at CHGC in Shanghai.

BASE COUNT 191 a 161 c 161 g 240 t  
 ORIGIN

Query Match 31.2%; Score 634.2; DB 32; Length 753;  
 Best Local Similarity 96.5%; Pred. No. 1.3e-145;  
 Matches 659; Conservative 0; Mismatches 23; Indels 1; Gaps 1;  
 OY 403 cctttagcttgacagtggtgcaattcattcattccttatttctgaggaacct 462  
 |||||  
 Db 25 CGTTTAAAGTCTGACAGTGTGCAATTCATTCATTCCTTATTTCGAGAAACCTC 84  
 |||||

OY 463 gtgtgttgacgtgtgtccacgttgaggaagagtgatattgtttaggaaggaacct 522  
 |||||  
 Db 85 GTGTGTTCAGTGTGTGCTCCAGTGAAGAAAGTGTATGTGTAGGAGCAACCTCA 144  
 |||||

OY 523 gtgttgagaaccttcagtcacgttgagcagctcgtaatgcctgtttccaagaana 582  
 |||||  
 Db 145 GTGTGTAAGACCTTTCAGTACCTTGGCCAGCTCCGTATATGCTTTCAAGAAAC 204  
 |||||

OY 583 tctgttccagttacccccccaattctctgtagttagaacaatgaagttgacctc 642  
 |||||  
 Db 205 TCTGTTCAGTACCTCCCTCCCAATTCCTCTGAGTATGAAACAATGAAGTTGACCTC 264  
 |||||

OY 643 ttcttttcgaacgtgcaagtgctacatgatatttcaagcttctgtctcgtcacaagat 702  
 |||||  
 Db 265 TTCTTTTGAACTGCAAGTGTCAACATGATTTTCAAGCTTGTCTGTGCAATTAACAT 324  
 |||||

OY 703 ctgacgaagatcatcttccttgattatatactacgtgagctggcaagtttgatgaatt 762  
 |||||  
 Db 325 CTACCAAGATCATTTCTCCGTATATATATATATATATATATATATATATATATAT 384  
 |||||

OY 763 gggaagcgttatggggaagacattgaaacattcagaatgctcttaagatcctgtgac 822  
 |||||  
 Db 385 GTTAAAGCTTATATGGAAGACTCTGCAACATTCATATAGATCTTCTATGACTCTG 444  
 |||||

OY 823 gctctgcaaaagtctgacagatgacatgacgtcttattgttggatgcagttgtaga 882  
 |||||  
 Db 445 GCTTCTCAAAATTTGCGATGATGACATGTACGTCTTTATGTGTGGAATGCGATGAG 504  
 |||||

OY 883 ttgttactgttaagatattgacacccctccatttagagaagaaggaactatccttga 942  
 |||||  
 Db 505 TTAGTACGTCAAGTCAATTTGACACCTCTCTTGTAGGAAGCAAGACATATCTTGA 564  
 |||||

OY 943 gcaaaacgaagcaagaccgaagctccctataaccttgacatataatatttga 1002  
 |||||  
 Db 565 GCAAAATCAGCAATTAACCATCAAGTCCCTATATCTTGCATATATATATATTTGA 624  
 |||||

OY 1003 tattcgtgttttcaacatgata-ctttgataatgatcgtgctgctgtgctgctg 1061  
 |||||  
 Db 625 TAATCGTGTGCTTCAACATGCGCCTATGATATATATATATATATATATATAT 684  
 |||||

OY 1062 tctacaccttcaacatatttga 1084  
 |||||  
 Db 685 TATACCTCTTCCCAATATTTGAA 707  
 |||||

RESULT 10  
 BP969051 954 bp mRNA EST 22-JAN-2001  
 LOCUS BP969051  
 DEFINITION 602269980P1 NIH\_MGC\_84 Homo sapiens cDNA clone IMAGE:4358275 5',  
 mRNA sequence.  
 ACCESSION BP969051  
 VERSION BP969051.1 GI:12336266  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 954)  
 AUTHORS NIH-MGC http://mgi.mcl.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: crabs-remail.nih.gov

Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L1AM995 row: p column: 20  
 High quality sequence stop: 652.  
 Location/Qualifiers  
 1..954

FEATURES  
 source  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4358275"  
 /clone\_lib="NIH MGC 84"  
 /tissue\_type="adrenal cortex carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site\_1:  
 NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dr  
 primed. Average insert size 1.229 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.

BASE COUNT 259 a 156 c 221 g 318 t  
 ORIGIN

Note: this is a NIH\_MGC library."

Query Match 31.1%; Score 633.2; DB 172; Length 954;  
 Best Local Similarity 92.8%; Pred. No. 2.4e-145;  
 Matches 687; Conservative 0; Mismatches 48; Indels 5; Gaps 2;

OY 992 atatttgaatattcgtgtgttttcaacatggtacttggataatgacgtcgttgcct 1051  
 DB 1 ATATTTTGAATATTCGTCGTTTAAATGATGATCTTGGATATGATCCCTTGGCC 60  
 OY 1052 tggctgtatatacaccctctacaatatttggaaatgacgtacccctgatatagaca 1111  
 DB 61 TGGCTGTGATATACCTCTTACAAATTTTGGAAATGATGATCCCTGATATATAGCATCA 120  
 OY 1112 ttatatagatacaaacaccagaagatcgaatggtatgtaatgtaacgttgcagaattag 1171  
 DB 121 TTTATATGATGATCAAAACCAAGATTCGAATGATGATGATGATGATGATGATGATG 180  
 OY 1172 aaaaaggggttggaaatctgctgttttgaataataatatttgaatcttgaatcttgaatga 1231  
 DB 181 AAAAGGGGTTGGAAATTTGGCTTTTGTAAATATATCTTTAGTGTGCTTTAAAGTA 240  
 OY 1232 gatagtatacttaccattataaaaaaaccaattgttgccttatttctgtgtgcc 1291  
 DB 241 GATAGTATATCTTTCATTTTAAAAAAATCAAAATTTTCTTATTTTGTGTGTGCC 300  
 OY 1292 tgtatgttttctcagaatgaattatgataatgacgtgaatccacgtgtgtatagatgc 1351  
 DB 301 TGTGATGTTTCTAGAGTGAATATATGATGATGATGATGATGATGATGATGATGATG 360  
 OY 1352 cataatagcttgaatattatgataagcaatataaacttgaatcttgaatcttgaatcttga 1411  
 DB 361 CATAATAGCTTGAATATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420  
 OY 1412 tgaatttgaataatgacgtgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 1471  
 DB 421 TGATTTTGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480  
 OY 1472 aaaaagctgactgaatattatgataagcaatataaacttgaatcttgaatcttgaatcttga 1531  
 DB 481 AAAAGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540  
 OY 1532 gttgaataatcatttgggtatctgtatataatgaatgaatgaatgaatgaatgaatgaatga 1591  
 DB 541 GGTGAATATCAATTTGGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
 OY 1592 tgtaaataacttgaataaagaagaagaagttttaaacttgaatgaatgaatgaatgaatga 1651  
 DB 601 TGTAA--TACTCTGAAACCAAGACAGAGTTTAACTTGAAGTACCTTAAATATTTGG 658  
 OY 1652 tgttctataataacgtctagtttggaaatgacgtatctgtaacgaagaagaagctgtt 1711  
 DB 659 ATGTGCTATATATGCGCTTAACTTGGAGCTGTCTGAG--TACGGAGAGGAGCTGTGTT 715  
 OY 1712 ttaaccctctctgcaagtt 1731  
 DB 716 TAACCCCTCTGCGCAGGTT 735

RESULT 11  
 BG403007 996 bp mRNA EST 12-MAR-2001  
 LOCUS 602418818F1 NIH\_MGC\_93 Homo sapiens cDNA clone IMAGE:4525781 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG403007  
 VERSION BG403007.1 GI:13296455  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
 1 (bases 1 to 996)  
 AUTHORS NIH-MGC <http://mhc.mci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [scaps-7@mail.nih.gov](mailto:scaps-7@mail.nih.gov)  
 Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLNL0432 row: d column: 06  
 High quality sequence stop: 702.  
 Location/Qualifiers

## FEATURES

1..996  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4525781"  
 /clone\_lib="NIH\_MGC\_93"  
 /tissue\_type="transitional cell papilloma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: bladder; Vector: pCMV-SPORT6; Site:1; Ncti:  
 Site:2; Salt: Cloned unidirectionally; oligo-dT primed.  
 Average insert size 1.7 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC library."

BASE COUNT 279 a 186 c 235 g 296 t  
 ORIGIN

Query Match 30.9%; Score 628; DB 153; Length 996;  
 Best Local Similarity 91.6%; Pred. No. 4.6e-144;  
 Matches 757; Conservative 0; Mismatches 50; Indels 19; Gaps 8;

OY 442 ttaatttctgggaagaactctgttcttctgcaatggtctcccaagtgaagaagaagtgtat 501  
 DB 1 TTTATTTCTGAGGAACCTCTGTGTTTTCAGATGATGATGATGATGATGATGATGATGATG 60  
 OY 502 atgttaggaaggaactcagtgatttgaagaacttgcagtcaccttgcgcagctccgt 561  
 DB 61 ATGTAGAGGAGGAACTCAGTCTTCTGAAAGACCTTTCAGTCACTTGCAGCAGCTCCGT 120  
 OY 562 aatgcgcgttctgaagaagaactctgtctcagttcactcccccactaattctcagtag 621  
 DB 121 AATGCGCTGTTTCAGAAACTCTGTTCTCAATGCAAGTCAATGCAATGCAATGCAATGCA 180  
 OY 622 aacaatgaagttagcctgtcttcttcttctgaactgaagtgtcatatattcaagc 681  
 DB 181 AACATGAAAGTTGACCTGCTCTTCTCTGAACTGCAAGTCAATGCAATGCAATGCAATGCA 240  
 OY 682 ttgctgtctgtaagaactcctgaagaagaatcctcctgattatattcaacggag 741  
 DB 241 TTGCTGTCTCTCTTAAGCATCTTACGCAAGATATTTCTCGATTATATTCAGTGGAG 300  
 OY 742 ctggcaagtttggatgaatctgggaagcgtatctgggaagactctgaacaattcaagat 801  
 DB 301 CTGGCAGGTTTGGATGAAATTTGGAAAGCTTATGGGAAGACTGCAATTCAGAGAT 360  
 OY 802 gcttctaagaatcctgttgaacgtctgcaaaagtctgagatgacatgataatcttatt 861  
 DB 361 GCTTCAAGATCTTGTGACGCTCTGCAAAAGTTTGCAATGCAATGCAATGCAATGCAATGCA 420  
 OY 862 gttggaatcagtgtagtagttagtcaatgcaatgcaatgcaatgcaatgcaatgcaatgca 921  
 DB 421 GGTGGAAATGCAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 480  
 OY 922 aagaagaactatccttggagcaaaagaagcaagcaagcaagcaagcaagcaagcaagcaagca 979  
 DB 481 AAGACAGAGATATCTTGAAGCAAAACAGCAAGACCAAGCAAGCAAGCAAGCAAGCAAGCAAG 540





```

OY 250 ttctctgtaaaagaaccttctctgcccaggacctcgagtggttaacctgttccatcgt 309
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 TTCTCTGTAAAGAGAACCTGTCTCTGCGAGGGCTTGCGGTGAACCTATCTCCACCG 300
OY 310 cctcgaggtaccgcatcagtgatgagtgaaaggagtgaaacaactggtctaaccccgagc 369
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 CCAAGGGGTACCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360
OY 370 agtgcattctgacccttggagaatggaatggaatggaatggaatggaatggaatggaat 429
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 AGGCTATCTCTTACCTTTTGAGAAATGCAATCTCTTTAGCTTACAGACGCGTGCATAAT 420
OY 430 tccattcactccttattcttgaggaactcctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 489
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 TCCATTCACTCTCTTTTTCGAGAGAAACACCTGAGTTTTCGAGTTGCGTCCACGAGAG 480
OY 490 gaaagagttatagtgatgaggaaggaactcagtggtgtgtgtgtgtgtgtgtgtgtgtgt 549
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 GAGAGAGTGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
OY 550 cgcacgctccgtaacgcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 609
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 541 CGGACACTCCGTAACCGCGCTGTTTCAAGAAACCTCCCTCTCAACTCTCTTCCCTCAAT 600
OY 610 tctctagtaggaagaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 669
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 601 TCTCTAGTAGGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGA 660
OY 670 gataattcgaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 729
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 661 GATATTTCAGATTGTTGCTGCTCATAGCATCATAGCATCATAGCATCATAGCATCATAG 720
OY 730 tattcagtagagctgagaggttgatgaggaatgaggaatgaggaatgaggaatgaggaat 789
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 721 TATCTATTGAGACTGCGAGGTTGGATGACTGCGGCTTCTCTCCGAACTGCAATGCTACAT 780
OY 790 caatcagagatgcttcaagatccttgtt 819
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 C-AGTCAGAGATGCTTNTAAGATCCTTGGT 809

```

RESULT 15  
 AM968992/c 648 bp mRNA EST 01-JUN-2000  
 LOCUS EST818169 MAGE resequences, MAGJ Homo sapiens cDNA, mRNA sequence.  
 DEFINITION AM968992  
 VERSION AM968992.1 GI:8158833  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 648)  
 AUTHORS Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt,  
 I. E., Saeed, A. I., Sharov, V., Lee, N. H., Yeatman, T. J. and  
 Quackenbush, J.  
 TITLE Assessment of gene expression patterns in a model of colon tumor  
 metastasis using a 19,200 element cDNA microarray  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: John Quackenbush  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 3528  
 Fax: 301 838 0208  
 Email: johnq@tigr.org  
 Plate: 259  
 Seq primer: Forward.  
 Location/Qualifiers  
 1. 648  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="MAGE resequences, MAGJ"  
 /note="Vector: pluescriptpskm"

```

BASE COUNT      226 a      124 c      86 g      212 t
ORIGIN
Query Match      29.6%; Score 601.2; DB 122; Length 648;
Best Local Similarity 98.7%; Pred. No. 1,7e-137;
Matches 606; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
OY 1083 gaacatgatacctcgtgataatgatacgaatcattatagatgatacgaacgaagatcgaat 1142
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 648 GAACCTGCTCCGATATGAAAGCCCTTTATAGCATGACCAACCGAAGATTCGAAT 589
OY 1143 ggaatgaatgatacctcgtgataatgatacgaatgatacgaatgatacgaatgatacgaat 1202
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 588 GGAATGAATGATTCCTGTCGAATATGAAAGGGGGTGGAAATGGCTGTTTGTGTA 529
OY 1203 aaatatacctttagtgctttaaagtagatagatatacattataaanaaat 1262
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 528 AATATATCTCTTATGATGCTTTAAAGTAGATAGTATACCTTATCATTTTAAAAAAAT 469
OY 1263 caattgtctccttattctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1322
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 468 CAATTTTGTCTTTATTTTGTGTGCGCTGTGATGTTTCTAGAGTAATTTATGATAT 409
OY 1323 tgaagtgaatccacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1382
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 408 TGACGTGAATCCACCTGCTGATAGATTCATTAATATGCTTGAATATATGATATAGCA 349
OY 1383 tttaataacattgattcattcctcgtttaaaggaatggaatggaatggaatggaatggaat 1442
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 348 TTTAATATACATTGATTTCATCTCTTTAATATGAAATTTGAAATATATGACTGAAAGAAATGT 289
OY 1443 aaacatttagaataagctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1502
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 288 AAAACATTTTAAATATGCTGCTGTATGGAAGAAAGTGCACATGAATTTTATGACCAAACTT 229
OY 1503 acaatgacttaacctcttacaacataggtgaaatcataattggcattgttact 1562
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 228 ACGAATGCTTACCTCTTTACACGACATAGGTGAATATCATTTTGGCTATTTGTACT 169
OY 1563 atgaacaatttgtaaatgctttaaatttgatgtaataactcgaacaagagaagaagt 1622
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 168 ATGAACATTTGTAAATGCTTTAATTTATGATTAATACCTGGAACAAAGAGAAAGGTT 109
OY 1623 ttttaactagatgagccctaaataatgagatgcttataataatgctttagtttgaact 1682
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 108 TTTAATCTAGAGTAGCCCTTAATAATATGATGCTTATATATCCCTTAGTTTGAAGT 49
OY 1683 gtactgagataca 1696
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 48 GTATCTGAGTACA 35

```

Search completed: August 18, 2001, 20:15:59  
 Job time: 7896 sec



---

